www.komputersapiens.org

# KOMPUTER SAPIENS

Revista de Divulgación de la Sociedad Mexicana de Inteligencia Artificial

Año 13
Volumen 3
Septiembre-Diciembre 2021

## SISTEMAS Y ALGORITMOS DE INTELIGENCIA ARTIFICIAL PARA COMBATIR LA VIOLENCIA DE GÉNERO

## RECUPERACIÓN DE MÚSICA BASADA EN ESTADOS EMOCIONALES: AVANCES ACTUALES Y RETOS FUTUROS

## IDENTIFICACIÓN AUTOMÁTICA Y ROBUSTA DE PARLANTES MEDIANTE REDES NEURONALES CONVOLUCIONALES

## TÉCNICAS DE MACHINE LEARNING PARA LA CLASIFICACIÓN DE ESTRÉS LABORAL: UN ESTUDIO COMPARATIVO

## IMPLEMENTACIÓN DE LA METODOLOGÍA CRISP-DM PARA EL PRONÓSTICO DE LOS PRECIOS DE LA UVA MEXICANA EN EL MERCADO DE ESTADOS UNIDOS

## INTELIGENCIA ARTIFICIAL EXPLICABLE XAI

Una exploración sobre el Aprendizaje Automático Simplificado: Generalización a partir de algunos ejemplos

SASTuit, software de análisis de sentimiento utilizando aprendizaje automático

Procesamiento de datos basado en homología persistente para identificación del COVID-19 en imágenes de rayos-x

Asistentes inteligentes para el manejo de datos geoespaciales ante desastres naturales

Aprendizaje de máquina en la clasificación y predicción del comportamiento de perros de búsqueda y rescate

ISSN 2007-0691

©Komputer Sapiens, Año XIII Volumen III, septiembre-diciembre 2021, es una publicación cuatrimestral de la Sociedad Mexicana de Inteligencia Artificial, A.C., con domicilio en Ezequiel Montes 56 s/n, Fracc. los Pilares, Metepec, Edo. de México, C.P. 52159, México, http://www.komputersapiens.org, correo electrónico: editorial@komputersapiens.org, tel. +52 (833)357.48.20 ext. 3024, fax +52 (833) 215.85.44. Impresa por Sistemas y Diseños de México S.A. de C.V., calle Aragón No. 190, colonia Álamos, delegación Benito Juárez, México D.F., C.P. 03400, México, se terminó de editar el 25 de septiembre de 2021.
Reserva de derechos al uso exclusivo número 04-2009-111110040200-102 otorgado por el Instituto Nacional de Derechos de Autor. ISSN 2007-0691.
Los artículos y columnas firmados son responsabilidad exclusiva de los autores y no reflejan necesariamente los puntos de vista de la Sociedad Mexicana de Inteligencia Artificial. La mención de empresas o productos específicos en las páginas de Komputer Sapiens no implica su respaldo por la Sociedad Mexicana de Inteligencia Artificial.
Queda estrictamente prohibida la reproducción total o parcial por cualquier medio, de la información aquí contenida sin autorización por escrito de los editores.
Komputer Sapiens es una revista de divulgación en idioma español de temas relacionados con la inteligencia artificial.
Creada en LaTeX, con la clase **papertex** disponible en el repositorio *CTAN*: Comprehensive TeX Archive Network, http://www.ctan.org/
**Indizada en el IRMDCT de CONACYT y en Latindex.**

**Directorio SMIA**

| | |
|---|---|
| Presidente | Félix A. Castro Espinoza |
| Vicepresidente | Ildar Batyrshin |
| Secretario | Oscar Herrera Alcántara |
| Tesorero | María de Lourdes G. Martínez Villaseñor |
| Vocales: | Sofía Natalia Galicia Haro |
| | Sabino Miranda Jiménez |
| | Roberto A. Vázquez Espinoza de los Monteros |
| | Hiram Erendin Ponce Espinosa |
| | Francisco Viveros-Jiménez |
| | Noé Alejandro Castro Sánchez |
| | Obdulia Pichardo Lagunas |
| | Omar Montaño Rivas |
| | Antonio Marín Hernández |
| | Néstor Velasco Bermeo |
| | Iris Iddaly Méndez Gurrola |
| | Gustavo Arroyo Figueroa |
| | Leobardo Morales Tiburcio |
| | Iván Olvera Rodríguez |
| | Juan Martínez Miranda |

**Komputer Sapiens**

| | |
|---|---|
| Director general | Félix A. Castro Espinoza |
| Editora en jefe | Karina Mariela Figueroa Mora |
| Editor científico | Héctor Gabriel Acosta Mesa |
| Editores asociados | Marco Antonio Aguirre Lam |
| | Laura Cruz Reyes |
| Editores invitados | Juana Canul-Reich |
| Coordinadora de producción | Viridiana Mena Gómez |
| Estado del IArte | Reyna Carolina Medina Ramírez |
| | Jorge Rafael Gutiérrez Pulido |
| Sakbe | Claudia G. Gómez Santillán |
| | Laura Cruz Reyes |
| | Laura Gómez Cruz |
| IA & Educación | María Lucía Barrón Estrada |
| | J. Julieta Noguez Monroy |
| | María Yasmín Hernández Pérez |
| Deskubriendo Konocimiento | Alejandro Guerra Hernández |
| | Leonardo Garrido Luna |
| Asistencia técnica | Alan G. Aguirre Lam |
| Corrección de estilo | Gilberto Rivera Zárate |
| | Marcela Quiroz Castellanos |
| | Ruth Esmeralda Barreda Guajardo |
| | Sonia Lilia Mestizo Gutiérrez |
| | Guillermo de Jesús Hoyos Rivera |
| | Susana Aurora Velasco Montiel |
| | Nicandro Cruz Ramírez |
| | Claudia Leticia Díaz González |
| | Leonardo Romero Muñoz |
| | Nelson Rangel Valdez |
| | José Antonio Camarena Ibarrola |
| | Luis Miguel García Velázquez |
| Edición de imagen | Laura Gómez Cruz |
| Portada | Magdaleno Granados Olvera, El Pez Gordo Design |

**Directores Fundadores**
Carlos Alberto Reyes García
Ángel Kuri Morales

**Comité Editorial**
Félix A. Castro Espinoza
Jesús Favela Vara
Sofía Natalia Galicia Haro
Miguel González Mendozao
Oscar Herrera Alcántara
Raúl Monroy Borja
Eduardo F. Morales Manzanares
Leonardo Garrido Luna
Carlos Alberto Reyes García
Angélica Muñoz Meléndez
Antonio Sánchez Aguilar
Luis Enrique Sucar Succar
Ángel Kuri Morales
José A. Martínez Flores
Juan Manuel Ahuactzin Larios
Manuel Montes y Gómez
Ofelia Cervantes Villagómez
Alexander Gelbukh
Grigori Sidorov
Laura Cruz Reyes
Ramon Brena Pinero
Juan Humberto Sossa Azuela

**Árbitros**
Ángel Kuri Morales
Carlos Barrios Hernández
César Cárdenas
Claudia Gómez
Daniela Alejandra Ochoa
David Ríos
Edgar Tello
Eduardo Morales
Efrén Mezura Montes
Federico Alonso Pecina
Giner Alor Hernández
Guillermo Hoyos Rivera
Guillermo Santamaría Bonfil
Héctor Avilés Arriaga
Héctor Rodríguez Rangel
David Terán Villanueva
Jorge Rodas Osollo
José Alberto Hernández Aguilar
José Francisco Martínez Trinidad
Karla Olmos Sánchez
Laura Cruz
Luis Castro
Matías Alvarado
Nancy Pérez Castro
Ramón Brena
Raúl Monroy Borja
Rocío Erandi Barrientos Martínez
Sara Garza Villarreal
Tania Turrubiates López

# Contenido

**Técnicas de aprendizaje automático para la clasificación de niveles de estrés laboral**
Pág. 7 ⇒Victor H. Benitez, Francisco M. Mockabee, Esthefania Ortega, Edna Baquera y Jesus Pacheco

**Identificación Automática y Robusta de Parlantes mediante Redes Neuronales Convolucionales**
Pág. 13 ⇒ José Antonio Camarena Ibarrola y Karina Mariela Figueroa Mora

**Sistemas y algoritmos de Inteligencia Artificial para combatir la violencia de género**
Pág. 19 ⇒ Obed Campos Solano, Pablo Pancardo y José Adán Hernández Nolasco

**Aprendizaje de máquina para la identificación del comportamiento de perros de búsqueda y rescate**
Pág. 24 ⇒ Rodrigo Chi Ozalde, Humberto Pérez Espinosa, Jorge Ríos Martinez

**Inteligencia Artificial Explicable XAI**
Pág. 29 ⇒ Tonantzin Marcayda Guerrero Velázquez y Juan Humberto Sossa Azuela

**Una Exploración Sobre el Aprendizaje Automático Simplificado: Generalización a partir de Algunos Ejemplos**
Pág. 36 ⇒ Alejandro Hernández Herrera, Elsa Rubio Espino, Rogelio Álvarez Vargas y Victor Hugo Ponce Ponce

**SASTuit, software de análisis de sentimiento utilizando aprendizaje automático**
Pág. 42 ⇒ Anibal A. Herrera Contreras, Eddy Sánchez De la Cruz, Ivan V. Meza Ruiz y Mirta Fuentes Ramos

**Panorama de la recuperación de música basada en estados emocionales**
Pág. 48 ⇒ Mónica Muñoz Ramírez, Humberto Pérez Espinosa, José Alberto Fernández Zepeda y Juan Martínez Miranda

**Análisis de técnicas inteligentes para la predicción del flujo de tráfico urbano**
Pág. 54 ⇒ Boris Medina Salgado, Eddy Sánchez De la Cruz, Pilar Pozos Parra y Javier E. Sierra

**Asistentes inteligentes para el manejo de datos geoespaciales ante desastres naturales**
Pág. 60 ⇒ Daniela Moctezuma, Julio Godoy y Diego Seco

**Implementación de la metodología CRISP-DM para el pronóstico de los precios de la uva sonorense en el mercado de Estados Unidos**
Pág. 65 ⇒ Vicente Solis Sandoval, Juan Martin Preciado Rodríguez y Luis Felipe Romero Dessens

## Columnas

Sapiens Piensa.
Editorial pág. 2

e-Tlakuilo pág. 4

Estado del IArte pág. 5

Sakbe pág. 6

IA & Educación pág. 70

Deskubriendo
Konocimiento pág. 72

# *Sapiens Piensa*

**Juana Canul-Reich**

Hoy en día diversos son los campos de aplicación de los algoritmos inteligentes. En este volumen de tema libre se logró la aceptación de once trabajos de investigación, los cuales dejan ver la diversidad de problemas que la comunidad científica está resolviendo por medio de algoritmos de la Inteligencia Artificial. Esta diversidad será aún mayor con el paso del tiempo, pues prácticamente todo problema tiene una solución computacional inteligente.

El artículo**Recuperación de música basada en estados emocionales: avances actuales y retos futuros** describe avances en el desarrollo de interfaces de búsqueda de música con base en criterios emocionales de los usuarios. Existen bases de datos que, además de los metadatos de cada una de las canciones, incluyen etiquetas alusivas al estado emocional representativo de la canción. Las etiquetas pueden representarse con modelos discretos o usando modelos dimensionales. Posteriormente, con el uso de algoritmos de aprendizaje automático como redes neuronales, máquinas de soporte vectorial (SVM, del inglés S*upport Vector Machine*) y más recientemente el aprendizaje profundo, es posible presentar resultados apropiados a la emoción que el usuario experimenta en ese momento. Con lo anterior, se observa que la investigación en este contexto podría dar lugar a una evolución en las plataformas de música en línea.

Por otro lado, en el artículo **Identificación Automática y Robusta de Parlantes mediante Redes Neuronales Convolucionales**, los autores presentan una técnica para el reconocimiento de individuos a través de su voz utilizando redes neuronales profundas, en particular las redes neuronales convolucionales (CNN, del inglés *Convolutional Neural Network*). Para ello las señales de voz se convierten en imágenes pequeñas de 24 x 32 pixeles llamadas entropigramas, bajo la hipótesis que éstas resultarían mejor que los espectogramas para señales ruidosas, lo cual logran comprobar con base en sus resultados experimentales.

Asimismo, el artículo **Implementación de la metodología CRISP-DM para el pronóstico de los precios de la uva mexicana en el mercado de Estados Unidos**, reporta la solución a un problema de pronóstico, una de las aplicaciones de la minería de datos. El problema es modelado como una serie de tiempo, con lo cual los autores pronostican el precio de la uva que el estado de Sonora exporta al mercado estadounidense para un periodo de 12 meses a futuro.

Los autores del artículo **Técnicas de Machine Learning para la clasificación de niveles de estrés laboral: Un estudio comparativo** aplican los algoritmos de redes neuronales, K Nearest Neighbor, árboles de decisión y SVM Lineal para crear modelos de detección de estrés laboral. Estos modelos son evaluados por medio de matrices de confusión y curvas ROC (del inglés *Receiver Operating Characteristic*). El mejor desempeño se obtiene con SVM, concluyéndose que estos modelos pueden servir de apoyo a los especialistas en salud laboral.

Por su parte, el artículo **Una Exploración Sobre el Aprendizaje Automático Simplificado: Generalización a Partir de Algunos Ejemplos** trata el concepto de entrenamiento de modelos de aprendizaje capaces de clasificar hacia nuevas categorías, utilizando muy pocos ejemplos y sin necesidad de un amplio reentrenamiento. Este trabajo representa una analogía a la capacidad de los seres humanos de reconocer sin ver ejemplos, y entra en contraste con las exitosas técnicas de aprendizaje profundo que requieren de gran cantidad de datos. Los autores describen las dos categorías de los modelos de aprendizaje simplificado y exponen algunas de sus aplicaciones.

Hoy día las redes sociales constituyen un campo de aplicación del procesamiento del lenguaje natural, dado que permite determinar las tendencias en el estado de ánimo de los usuarios en relación con un tema específico. En el artículo **SASTuit, software de análisis de sentimiento utilizando aprendizaje automático**, los autores presentan la herramienta SASTuit que está enfocada al análisis de tweets en español, y es capaz de clasificar el sentimiento vertido en los tweets en positivos y negativos.

El artículo **Inteligencia Artificial Explicable XAI** describe que esta área consiste en mejorar el entendimiento humano sobre los modelos artificiales. La investigación realizada propone que la principal desventaja de los métodos de explicación visuales es la subjetividad de sus resultados. Los autores contribuyen al área con un método útil para clasificadores de imágenes que ofrece explicaciones no subjetivas, el cual permite que la interpretación del resultado y su explicación siempre sea la misma.

Otra aplicación de los algoritmos inteligentes está en el procesamiento del tráfico urbano estudiado en el artículo **Análisis de técnicas inteligentes para la predicción de flujo de tráfico urbano**. En este trabajo los autores analizan diferentes técnicas encontradas a partir de una revisión de trabajos de investigación relacionados que tienen una antigüedad no mayor a cinco años.

El artículo organiza las técnicas dentro de una taxonomía compuesta por tres categorías, las paramétricas,

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

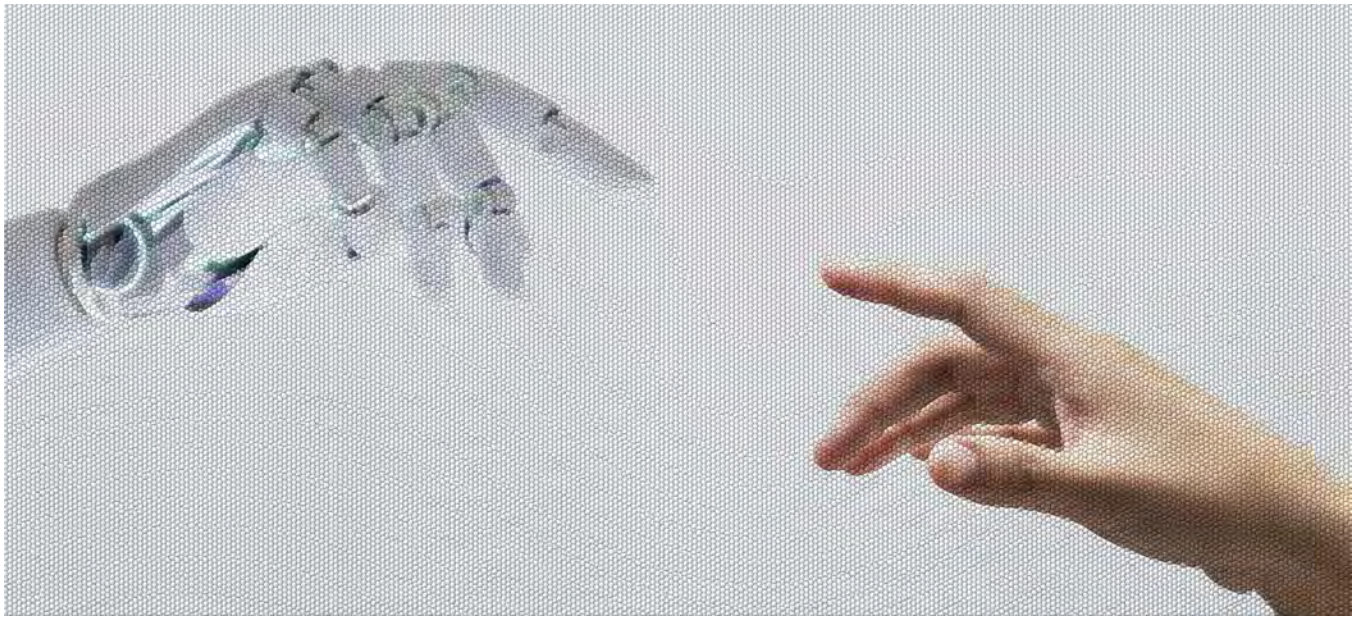
las de aprendizaje automático y las de aprendizaje profundo.

Del mismo modo, el artículo **Asistentes inteligentes para el manejo de datos geoespaciales ante desastres naturales** presenta los avances actuales en el desarrollo de asistentes inteligentes, expone los principales retos para su uso en situaciones de desastres naturales y puntualiza en los elementos esenciales ante estas situaciones. Al final, se presenta el diseño conceptual del asistente propuesto en la red IDEAIS.

Una aplicación adicional de algoritmos inteligentes es en situaciones de desastre, como se reporta en el artículo **Aprendizaje de máquina en la clasificación y predicción del comportamiento de perros de búsqueda y rescate**. Este artículo propone un método que, haciendo uso de dispositivos electrónicos vestibles, captura información auditiva, de movimiento y biométrica de perros, las cuales al combinarlas permiten determinar su estado interno, sus emociones y la actividad que realizan. El objetivo final es perfilar a los perros de búsqueda y rescate en su entrenamiento.

Por último, el artículo **Sistemas y algoritmos de inteligencia artificial para combatir la violencia de género**, presenta un estudio en dos partes. Primero, hace un recuento de sistemas y aplicaciones para detectar la violencia de género. Los sistemas descritos hacen uso de objetos tales como anillos inteligentes, pulseras GPS, brazaletes entre otros para activar alguna funcionalidad de alerta. Segundo, hace un recuento de algoritmos de Inteligencia Artificial para detectar violencia de género. Entre los algoritmos aplicados se encuentran Máquinas de Soporte Vectorial, Redes Neuronales Convolucionales, Redes Bayesianas, entre otros.✳

**Dra. Juana Canul-Reich** es Profesora-Investigadora en la División Académica de Ciencias y Tecnologías de la Información de la Universidad Juárez Autónoma de Tabasco (UJAT-DACYTI), México. Sus intereses en investigación se enfocan principalmente en la Ciencia de Datos e Inteligencia Artificial para resolver problemáticas en el área de la Salud.



# e-Tlakuilo: Cartas de nuestros lectores

Rosa María Valdovinos
etlakuilo@komputersapiens.org

En *Komputer Sapiens* nos hemos esforzado por estar "a sólo un click de distancia" a través de diferentes medios como Facebook, Twitter y correo electrónico. Les presentamos algunas de las preguntas que hemos recibido a través de estos medios.

**Anónimo – Estudiante de nivel Posgrado. (vía correo electrónico)**
En el número más reciente, se publicó el artículo "Automatización y control difuso de un tractor para su servicio en agricultura", mi pregunta a los autores es la siguiente: ¿Por qué se necesitó un sistema de posicionamiento RTK-GPS para lograr una precisión de 1cm si el sistema cuenta con un sistema de reconocimiento de imágenes? Es decir, se podría colocar en un límite de la parcela y, con lo que describe el artículo, debería ser capaz de determinar los demás límites así como la forma de navegar dentro de ella a través de la detección de surcos mediante *clustering*.

**El autor, Marco Antonio Márquez Vera, responde:**

Con frecuencia se emplean sistemas redundantes para controlar la navegación de vehículos autónomos. En el caso analizado se pretende que no sólo sea redundante la manera de retroalimentar la información que llega al tractor, sino también que se complemente con otros sistemas. Empleando únicamente visión artificial sería posible navegar de un punto a otro en la parcela, teniendo dificultad en seguir una trayectoria fija debido a las irregularidades del terreno y a los obstáculos con que se encuentre el tractor; una alternativa sería determinar puntos visibles para la cámara cada cierta distancia, situación que parece ser resuelta empleando un RTK-GPS con el que es posible determinar la posición del tractor con una resolución de 1cm. El método anterior se aplica ya en algunas marcas de tractores comerciales, como se mencionó en el artículo; con él existe la posibilidad de determinar la trayectoria que seguirá el tractor para hacer o trasladarse por los surcos. Desafortunadamente, para que esto sea posible es necesario haber tratado el terreno, antes de que el tractor pueda hacer su tarea. Gracias al sistema redundante, el GPS indicará la posición actual del tractor, para que navegue de manera autónoma en la parcela, y ante la presencia de obstáculos (no detectados por el GPS tales como piedras, arbustos, o surcos no lineales) la visión artificial tomaría el control del tractor para modificar la trayectoria con el fin de esquivarlos, y con la ayuda nuevamente del GPS se buscará regresarlo a la ruta antes fijada.

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

# Estado del IArte

Jorge Rafael Gutiérrez Pulido (@jrgpulido) y Reyna Carolina Medina Ramírez
estadoiarte@komputersapiens.org

## Chatbots Educativos

Los chatbots, conocidos también como asistentes virtuales o Bots conversacionales, son programas que integran diferentes tecnologías y técnicas específicas inherentes a campos de conocimiento de la inteligencia artificial. En el desarrollo de los chatbots se involucran por ejemplo el Procesamiento del Lenguaje Natural (PLN o NLP), los árboles de decisión y el aprendizaje de máquina, entre otros.

Los chatbots están orientados a problemas concretos y son capaces de simular o mantener cierto nivel de conversación con las personas, a través de un guion conversacional utilizando respuestas predeterminadas en un contexto específico dando una sensación de inmediatez. La conversación entre el usuario y el Bot puede ser de tipo voz o texto empleando el PLN para procesar el guion conversacional. En la línea de tiempo de los chatbots tenemos desde Eliza en 1966 hasta SIRI (2010), Cortana (2014), Alexa (2015), Google Now (2016), Bixby (2017) por mencionar solo algunos.

Dentro de las aplicaciones en el entorno educativo podemos encontrar los chatbots de apoyo al alumno que se enfocan en orientarlo en trámites administrativos o académicos por ejemplo planes de estudio, procedimiento de titulación, servicios de la institución educativa. Por otro lado, tenemos los chatbots de apoyo al aprendizaje o Bots de enseñanza personalizada, entre los que destacan los chatbots mentores o tutores, los de práctica y los de evaluación. Todos ellos enfocados al análisis de las necesidades de cada alumno y sirviendo de apoyo a docentes en la enseñanza, asesoría o práctica. Algunas de las ventajas de utilizar chatbots educativos en el proceso de enseñanza-aprendizaje son: conversación amigable con los alumnos; servir de guía en el aprendizaje de un tema, respondiendo dudas ya sea del tipo teórico o práctico; disponibilidad a través de aulas virtuales del tipo Classroom, Teams, Moodle, sitios web institucionales o redes sociales; automatizar tareas específicas, responder preguntas frecuentes, flexibilidad para aprender a un ritmo acorde al establecido por el estudiante, y adaptabilidad a situaciones particulares de los alumnos. A continuación presentamos algunos chatbots destacados.

**Jill** es una asistente de enseñanza creada sobre la plataforma Watson de IBM. En la Universidad de Georgia, Jill proporciona apoyo a estudiantes de nivel posgrado en el curso de Inteligencia Artificial basada en el conocimiento (KBAI), teniendo una disponibilidad de 24/7. Para conocer más `https://b.gatech.edu/37pgzOT`

**Genie** (disponible también en Google Play) permite a los estudiantes de la Universidad de Deakin acceder fácilmente a sus horarios, resultados, información de la Universidad, así como a una variedad de respuestas a preguntas comunes que plantean otros estudiantes. Este chatbot asiste a los estudiantes a través del registro de sus necesidades y les manda notificaciones como: tienes estas actividades pendientes, te recuerdo examen de matemáticas estudiar esta noche. Para conocer más `https://bit.ly/37t1VGE`

**MOOCBot** es un chatbot con características de un Bot mentor o tutor y un Bot evaluador cuyo objetivo es ayudar a los estudiantes de la asignatura Innovaciones en e-learning del Master en Educación y Nuevas Tecnologías a la elaboración del diseño instruccional de un MOOC y en la evaluación de las propuestas realizadas por los compañeros de clase. Para conocer más `https://bit.ly/3ywC4tl`

**EducaBots** es una trilogía de Bots en las categorías de tutores, de práctica y evaluadores (TeoBot, PractiBot y QuizBot) sobre temas de bases de datos de la licenciatura en Computación en la UAM-I. El objetivo de estos Bots es servir de apoyo tanto alumnos, como a profesores en el proceso de enseñanza-aprendizaje de temas teóricos y prácticos, así como poder medir la adquisición de estos. Una primera versión de esta trilogía se puede explorar en `https://bit.ly/3CtXI3N`, `https://bit.ly/3s0LKd2`, y `https://bit.ly/3yGcGS0`✷

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

# Sakbe

**Laura Gómez Cruz**, **Laura Cruz Reyes** y **Claudia Guadalupe Gómez Santillán**
sakbe@komputersapiens.org

## Mujeres y algoritmos de inteligencia artificial para combatir la violencia de género

Hablar de violencia de género es un tema complejo, porque abarca desde amenazas económicas y psicológicas hasta agresiones físicas y sexuales a individuos o grupos de personas, debido a la desigualdad de poder de su género. En las mujeres, esta desigualdad la podemos percibir en el trato familiar, laboral y profesional, que propicia desvalorización y pérdida de confianza, así como una banalización de la violencia contra ellas. Por la razón anterior, el desarrollo de algoritmos y tecnologías imparciales de inteligencia artificial (IA) para prevenir la violencia es de sumo interés en la actualidad. Otra estrategia relevante, pero poco considerada, es corregir la escasa participación de la mujer en la IA para evitar que las máquinas se vuelvan machistas.

En su sitio web, la ONU presenta iniciativas con la intención de romper el sesgo de género en la IA. Una de las grandes preocupaciones es que los desequilibrios de género están profundamente arraigados y pueden distorsionar el desarrollo de la IA, para hacer retroceder varias décadas de lucha por la igualdad de género. Los siguientes datos son una muestra que revelan este desequilibrio en contra de las mujeres a nivel mundial: sólo representan el 22 % de los profesionales de la IA, tienen cuatro veces menos probabilidades de tener conocimientos de programación informática, tienen menor probabilidad de acceder a puestos directivos en el área y enfrentan el riesgo de propagar estereotipos de sumisión y servilismo, por ejemplo, a través a las IA de asistencia que actualmente tienen voz femenina.

La ONU y el Foro Económico Mundial organizaron una mesa redonda sobre la igualdad de género y las mujeres en la IA. Se abordaron tópicos como: sensibilización sobre los prejuicios sexistas en la IA, aprendizaje automático para el cumplimiento de leyes sobre la esclavitud moderna, algoritmos inclusivos, reducción de prejuicios en twitter y ética para la IA.✳

Para saber más sobre cómo puede ayudarnos la IA a combatir la violencia de género consulta los siguientes sitios Web:

- **Mesa redonda sobre la igualdad de género y las mujeres en la IA**, incluye video del evento:
  - `https://www.onu.org.mx/mujeres-e-inteligencia-artificial/`
  - `https://en.unesco.org/girltrouble`
- **La ausencia de mujeres en el campo de la inteligencia artificial reproduce el sexismo:**
  - `https://news.un.org/es/story/2019/06/1456961`
- **La inteligencia artificial ayuda a prevenir la violencia de género:**
  - `https://elpais.com/tecnologia/haz-cosas-extraordinarias/2021-02-18/la-inteligencia-artificial-ayuda-a-prevenir-la-violencia-de-genero.html`
- **Algoritmos contra la violencia machista:**
  - `https://www.lavanguardia.com/tecnologia/20190519/462147339117/viogen-violencia-de-genero-violencia-machista-inteligencia-artificial-algoritmos.html`

Combate a la desigualdad de genero con la IA

ARTÍCULO ACEPTADO

# Técnicas de aprendizaje automático para la clasificación de niveles de estrés laboral: Un estudio comparativo

Victor H. Benitez, Francisco M. Mockabee, Esthefania Ortega, Edna Baquera y Jesus Pacheco

El estrés laboral es un fenómeno mundial que genera síntomas de estrés general además del desgaste profesional, caracterizado por una falta de interés, negativismo hacia el trabajo y baja productividad. Múltiples gobiernos alrededor del mundo han buscado implementar normas para la regularización del entorno de trabajo, mejora de condiciones de los trabajadores, así como prevención y corrección del estrés laboral. En México, se aprobó la norma NOM-035-STPS-2018 (NOM-035), con el objetivo de identificar factores de riesgo psicosocial que puedan provocar estrés laboral. Existen diversos métodos para la detección de estrés; sin embargo, estos suelen ser costosos o tardados, por lo que el presente artículo propone un método de detección de estrés laboral utilizando modelos de aprendizaje automático entrenados a partir de un conjunto de datos generado al aplicar un instrumento de medición, siguiendo la NOM-035. Se probaron dos modelos: redes neuronales y árboles de decisión, con la intención de comparar su desempeño ante la problemática planteada.

El estudio presentado en este trabajo aborda la problemática de usar algoritmos de aprendizaje automático para una medición real llevada a cabo en una empresa desarrolladora de software. El objetivo es explorar el aprendizaje automático para apoyar las decisiones que deben tomar los encargados de la salud de los trabajadores y cumplir la citada norma oficial.

## Introducción

El estrés es un fenómeno psicosocial que ha sido motivo de preocupación a nivel mundial puesto que se ha reportado como una condición persistente en los centros de trabajo, provocada por diversos factores, y teniendo repercusiones negativas tanto en la salud física y mental como en el rendimiento, pudiendo causar también deterioro cognitivo en el individuo [1]. En este documento se presentan dos modelos de inteligencia artificial para la clasificación del nivel de estrés en el ámbito laboral, esto con el fin de ayudar a la regularización del entorno laboral según la NOM-035, la cual tiene como objetivo "Establecer los elementos para identificar, analizar y prevenir los factores de riesgo psicosocial, así como para promover un entorno organizacional favorable en los centros de trabajo" [2] (p. 3). El desarrollo de los modelos se realizó en MATLAB, el cual facilita la creación de diversos modelos de inteligencia artificial como son las redes neuronales y los árboles de decisión.

## Antecedentes

El estrés se puede definir como todo aquel estímulo que provoca tensión física, emocional o psicológica [3] y puede ser categorizado en tres tipos: agudo, agudo episódico y crónico. Existen diversos estudios que comprueban que el sometimiento prolongado a altos niveles de estrés puede provocar atrofia de masa cerebral [3], enfermedades cardiovasculares, cáncer, Alzhéimer, depresión, ansiedad, obesidad y diabetes, entre otras [4]. El estrés en el entorno de trabajo se denomina "estrés laboral" y puede ser provocado por condiciones físicas como temperatura, ruido o luminosidad, así como horarios excesivos, una gran carga laboral, nivel ocupacional o relaciones sociales en el entorno laboral [5]. Los efectos del estrés laboral se caracterizan por una falta de interés, negativismo hacia el trabajo y baja productividad.

En México, se aprobó la norma NOM-035-STPS-2018 "Factores de riesgo psicosocial en el trabajo – Identificación, análisis y prevención" [2], donde se propone una herramienta que provee una lista de variables indicadoras para la detección de empleados propensos al estrés laboral causado por riesgos psicosociales. Evidencias de su aplicación en diferentes centros de trabajo demuestran que ésta es una herramienta útil que provee resultados satisfactorios para la detección temprana de factores que permiten establecer estrategias de acción para la reducción de riesgo psicosocial [6, 7].

Existen además otros métodos para el diagnóstico de estrés, como lo son pruebas hormonales y electrocardiogramas [4]; sin embargo, éstas pueden ser costosas y tardadas, por lo que se han propuesto diversos métodos para una detección rápida, eficaz y accesible, dentro de los que se encuentra la aplicación de técnicas de aprendizaje automático para la identificación de sujetos estresados a partir de publicaciones en redes sociales, el procesamiento de resultados a sensores rápidos que monitorean tensión muscular, temperatura corporal, ritmo cardiaco, entre otros [8]. En el presente trabajo, se busca obtener un modelo funcional para la rápida detección de riesgo de estrés laboral acorde a la NOM-035, utilizando los siguientes modelos: redes neuronales [9] y árboles de decisión [10].

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

## Metodología

Para el diseño de los modelos, se utilizó un conjunto de datos de 221 muestras con 75 variables, las cuales se categorizan en tres clases: "estresado", "semi-estresado" y "no estresado", contando con 7, 15 y 199 muestras de cada clase respectivamente. Los rangos de las variables son:

$$nivelEstres = \begin{cases} \text{no estresado} & \text{si } x < 100, \\ \text{semi-estresado} & \text{si } 100 \leq x \leq 120, \\ \text{estresado} & \text{en otro caso.} \end{cases}$$

El procesamiento de datos se realizó utilizando Microsoft Excel y MATLAB como soporte.

## Diseño del modelo neuronal

En la Figura 1 se muestra la arquitectura de la red neuronal diseñada para esta aplicación.

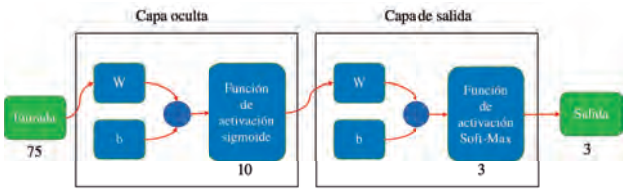



**Figura 1.** Arquitectura de la red neuronal.

*Entradas:* Debido a la cantidad de variables, la red cuenta con 75 entradas, con los cuales aprenderá a clasificar las muestras en las categorías "estresado", "semi-estresado" y "no estresado".

*Capa oculta:* La red cuenta con 10 neuronas en su capa oculta, dicho número de neuronas se eligió de forma heurística al verificar varias veces el desempeño de la red con diferentes cantidades de neuronas.

*Función de activación de la capa oculta:* La capa oculta utiliza una función de tipo sigmoidal hiperbólica para la activación de las neuronas, por lo que las salidas de la capa oculta varían en un rango de -1 a +1.

*Capa de salida:* Esta capa cuenta con solo tres neuronas, esto debido a que existen tres clases posibles para clasificar las muestras.

*Función de activación de la capa de salida:* La capa de salida utiliza una función de activación SoftMax, que se caracteriza por calcular la probabilidad de pertenencia que tiene una muestra hacia cada una de las clases.

*Salida:* Habiendo tres neuronas en la capa de salida, habrá tres salidas en la red neuronal, correspondientes a las probabilidades de pertenencia de la muestra analizada para cada categoría.

*Algoritmo de aprendizaje:* Se utilizó el algoritmo de retropropagación. El conjunto de datos fue dividido de la siguiente forma: 70 % para entrenamiento, 15 % para validación y 15 % para pruebas. El desempeño durante el entrenamiento para cada subconjunto de datos se muestra en la Figura 2.

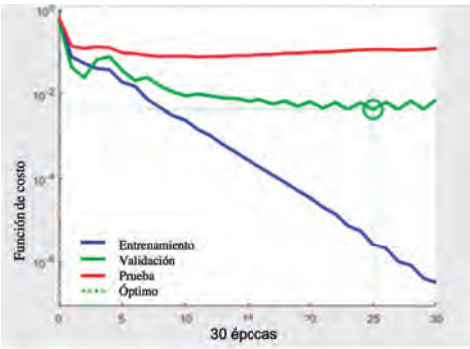



**Figura 2.** Desempeño del entrenamiento.

En las Figuras 3–6 se muestran las matrices de confusión, las cuales permiten visualizar el desempeño del método utilizado al clasificar.

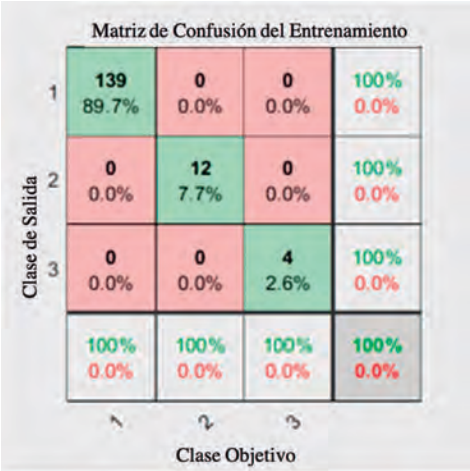



**Figura 3.** Matriz de confusión del entrenamiento.

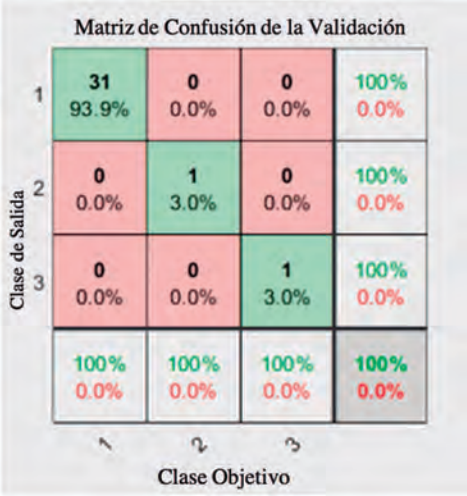



**Figura 4.** Matriz de confusión de la validación.

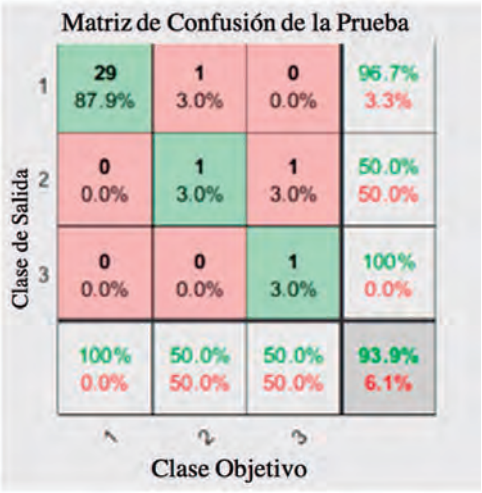



**Figura 5.** Matriz de confusión de la prueba.

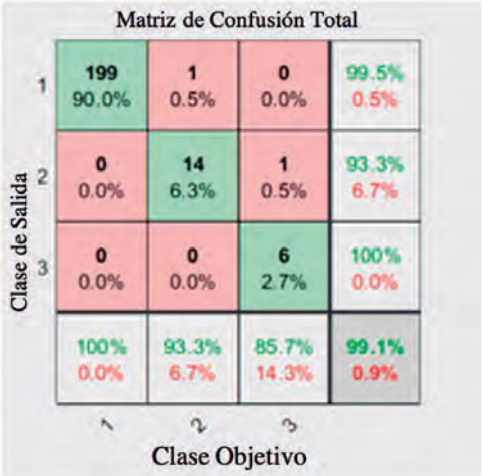



**Figura 6.** Matriz de confusión total.

## Interpretación de resultados de la red neuronal

Luego de entrenar varias veces la red, observar los resultados del entrenamiento y sus métricas (Figuras 3–6) y analizar el *dataset* utilizado, se observa lo siguiente:

Debido a los pocos casos de la clase "estresado" y "semi-estresado" con los que cuenta el conjunto de entrenamiento, se presentan irregularidades al efectuar el entrenamiento, ya que por la forma en la que se divide el conjunto de datos entre entrenamiento (70 %), prueba (15 %) y validación (15 %), hay una considerable probabilidad de que el conjunto de entrenamiento no contenga muestras de alguna de las clases, por lo que la red nunca aprendería de dichos ejemplos. Por otro lado, en el caso inverso; es decir, en el que los casos de las clases "estresado" o "semi-estresado" quedaran en el conjunto de entrenamiento, esto dejaría sin casos para comprobar la efectividad de la red con los conjuntos de validación y prueba, por lo que no habría forma de saber si el modelo clasifica correctamente las muestras. Suponiendo un caso ideal en el cual cada partición del conjunto tuviera una de las tres muestras de las clases "estresado" y "semi-estresado", esto no supondría una mejora, ya que, habiendo una cantidad pequeña de muestras al compararse con la cantidad de muestras de la clase "no estresado", esto no sería suficiente para entrenar la red. Este problema tampoco se puede solucionar reentrenando la red ya que se podría caer en sobreajuste, e incluso, también hay una alta probabilidad de tener sub-ajuste, debido a la cantidad de muestras de la clase "estresado".

Los resultados obtenidos para la red neuronal abren dos posibilidades de mejora: a) mezclar de forma aleatoria el conjunto de datos antes de dividirlo en conjuntos de entrenamiento, prueba y validación, esperando que de esa manera se tenga una mayor probabilidad de tener todas las clases bien representadas en los subconjuntos que se le mostrarán a la red neuronal o, b) lo ideal sería seguir ampliando el conjunto de datos hasta que se tenga una cantidad considerable de casos que le permitan a la red neuronal entrenarse de forma adecuada.

## Modelo tipo árbol de decisión

Se diseñó un árbol simple de clasificación y regresión con un número máximo de nodos internos igual a cuatro e índice Gini como criterio de división. Se usó una validación cruzada de orden 5 para garantizar independencia de resultados. El resultado se muestra en la Figura 7, donde VAR31, VAR14, etc., indican el número de variable del conjunto de datos.

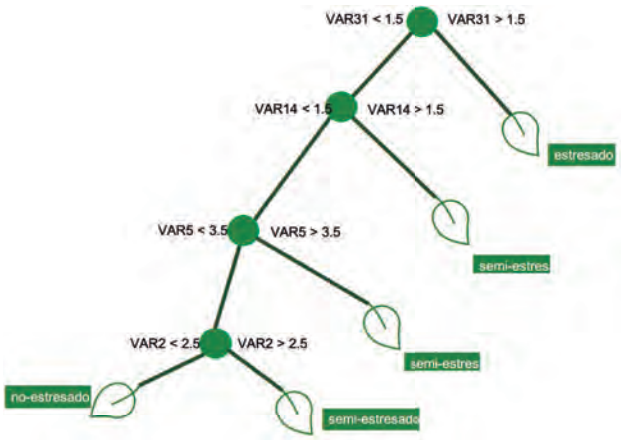



**Figura 7.** Modelo de árbol de decisiones para el conjunto de datos de estrés laboral.

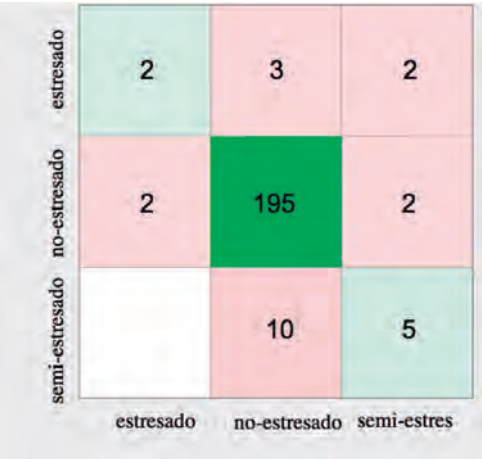



**Figura 8.** Matriz de confusión del árbol de decisión.

En la matriz de confusión de la Figura 9, se puede observar que la clasificación de la clase "estresado" fue correcta en 2 ocasiones y falló en otras 5, la clasificación de "no estresado" fue acertada en 195 ocasiones y falló en un total de 4, y para los casos de "semi-estresado" se clasificó correctamente en 5 casos y falló en un total de 10.

De acuerdo con la matriz de confusión mostrada en la Figura 8, se observa que el modelo clasifica de manera correcta el 30 % de las muestras correspondientes a los casos de "estresado"; 33 % para "semi-estresado" y 98 % para "no estresado". El rendimiento del árbol entrenado es evaluado mediante las curvas características operativas del receptor (ROC) para cada clase, donde se muestra la sensitividad del modelo contra la tasa de falsas alarmas para las tres clases, se reporta como el caso de la clase "no estresado" en la Figura 9 para ilustrar la técnica.

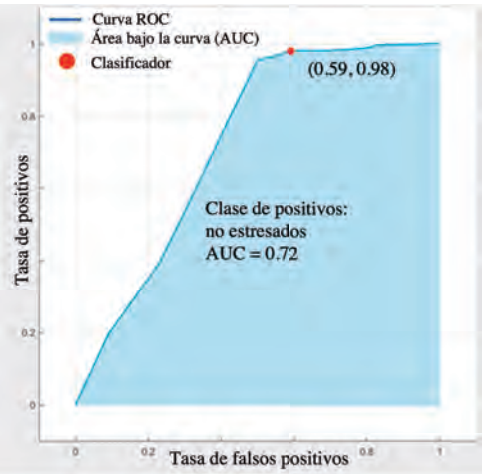



**Figura 9.** Gráfico ROC para los casos de "no estresado".

## Interpretación de resultados para el árbol de decisión

Es posible establecer que el modelo muestra un desempeño aceptable, pese a que presenta una precisión del 98 % para una de las clases. La diferencia en el número de muestras jugó un papel importante de este resultado, ya que se tiene una proporción muy grande de casos de "no estresado" en comparación con la suma de las otras dos clases. Aunado a la cantidad de datos, la gráfica ROC es un fuerte indicador del comportamiento del modelo. Para la clase "no estresado" la curva sube de forma vertical en el lado izquierdo con un área de 0.72; mientras que para la clase "estresado", el área bajo la curva es de 0.66, lo cual es un buen valor si se considera la baja representatividad de muestras en el dataset para esta clase. Para el caso "semi-estresado", el área bajo la curva estuvo alrededor de 0.7 por lo que se puede considerar un modelo útil, pero aun así es un porcentaje bajo para identificación de casos, especialmente si son de la clase "estresado".

## Conclusiones y discusión de resultados

El tamaño del conjunto de datos al estar desbalanceado en las muestras, afectó en diferente medida en los dos modelos de aprendizaje automático usados. En el caso de la red neuronal, se observa que el modelo estaría siempre en riesgo de caer en sub-ajuste o sobreajuste; los resultados en el entrenamiento fueron buenos, pero funcionó de manera pobre para la validación y la prueba. Parte del problema podría deberse a la distribución de los datos de "estresado" en las pruebas.

En el caso del árbol de decisión, se produjo una precisión de 87 %. El área bajo la curva es de alrededor del 70 %, lo cual representa resultados prometedores dado que los árboles son más fáciles de implementar.

Finalmente, con los resultados obtenidos en este estudio indagatorio, poniendo a prueba dos modelos populares de aprendizaje automático, se puede establecer que los resultados son bastante alentadores. Los modelos diseñados son simples y susceptibles de mejoras, tales como sintonizar parámetros óptimos para la red neuronal o utilizar árboles de decisión más complejos. Es necesario concluir que lo más importante es seguir tomando mediciones e incrementar la cantidad de muestras de las clases del conjunto de datos para obtener una representatividad más proporcionada de las clases.

Los hallazgos de este trabajo indican que es posible usar un algoritmo de aprendizaje automático que sirva de apoyo a los especialistas en salud laboral a tomar decisiones respecto a programas que les permitan aplicar de mejor forma la norma NOM-035-STPS-2018. De tal manera que los responsables de tomar decisiones de tratamiento y mitigación del estrés laboral puedan apoyarse en una herramienta de inteligencia artificial como las mostradas para aplicar procedimientos más precisos con

las poblaciones identificadas previamente por el modelo entrenado. Como trabajo futuro se busca implementar en tiempo real una herramienta de apoyo para las empresas basadas en alguno de los modelos estudiados dada su fácil implementación y codificación existente en la actualidad. ✷

## REFERENCIAS


1. Ananya, B., Priya, V., y Gayathri, R. (2019). Stress and health problems among information technology professionals. *Drug Invention Today*, 12(7), 1336-1338.

2. Secretaría de Trabajo y Previsión Social. (2018). Norma Oficial Mexicana NOM-035-STPS-2018, Factores de riesgo psicosocial en el trabajo-Identificación, análisis y prevención. Diario Oficial de la Federación.

3. Yaribeygi, H., Panahi, Y., Sahraei, H., Johnston, T. P., y Sahebkar, A. (2017). The impact of stress on body function: A review. *EXCLI journal*, 16, 1057–1072.

4. Devi, P., Reddy, A., Zahan, O., y Sharma, J. (2019). The effect of stress on human life. *Adalya Journal*, 8(9), 792–811.

5. Coduti, P., Gattás, Y., Sarmiento, S., y Schmid, R. (2013). Enfermedades laborales: cómo afectan el entorno organizacional. Tesis de licenciatura. Licenciatura en Contaduría Pública. Universidad Nacional de Cuyo.

6. Alaniz, J., y López, L. (2021). Implementación del diagnóstico de factores de riesgo psicosocial con base a la NOM-035-STPS-2018 a personal operativo de una organización educativa. *Revista de Desarrollo Sustentable, Negocios, Emprendimiento y Educación*, 17, 94–108.

7. Ortega, T. (2020). Diagnóstico Organizacional de Aerolíneas Ejecutivas S.A. de C.V.: Evaluación de la Cultura y el Clima Laboral e Identificación de los Factores de Riesgo Psicosociales con base en la NOM-035-STPS-2018. Tesis de licenciatura, Licenciatura en Comunicación. Universidad Autónoma del Estado de México.

8. Baheti, R., y Kinariwala, S. (2019). Detection and Analysis of Stress using Machine Learning Techniques. *International Journal of Engineering and Advanced Technology*, 9(1), 335–342.

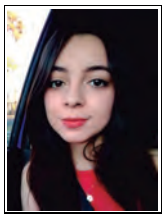
9. Mohamad, H. (2021). Fundamentals of Artificial Neural Networks. *MIT Press. A Bradford Book.* ISBN 978-0262514675.

10. Wlodarczak, P. (2019). Machine Learning and its Applications (1st ed.). *CRC Press.* ISBN 978-1138328228.


## SOBRE LOS AUTORES

**Victor H. Benitez** Se graduó como Ingeniero en Comunicaciones y Electrónica por la Universidad de Guadalajara. Obtuvo la Maestría en Ciencias y el Doctorado en Ciencias con Especialidad en Ingeniería Eléctrica en el CINVESTAV - IPN. Ha trabajado en la industria de la manufactura de circuitos impresos y en ingeniería de pruebas. Actualmente es profesor titular de Ingeniería Mecatrónica en la Universidad de Sonora y profesor de cátedra en el Tecnológico de Monterrey. Sus intereses de investigación incluyen el diseño de sistemas de control mioeléctrico y aplicaciones de sistemas inteligentes con un enfoque multidisciplinario.

**Francisco Mockabee** es estudiante de ingeniería mecatrónica en la Universidad de Sonora y miembro de la IEEE Computer Society rama Búhos, sus áreas de interés son la ingeniería de control, la inteligencia artificial y la robótica

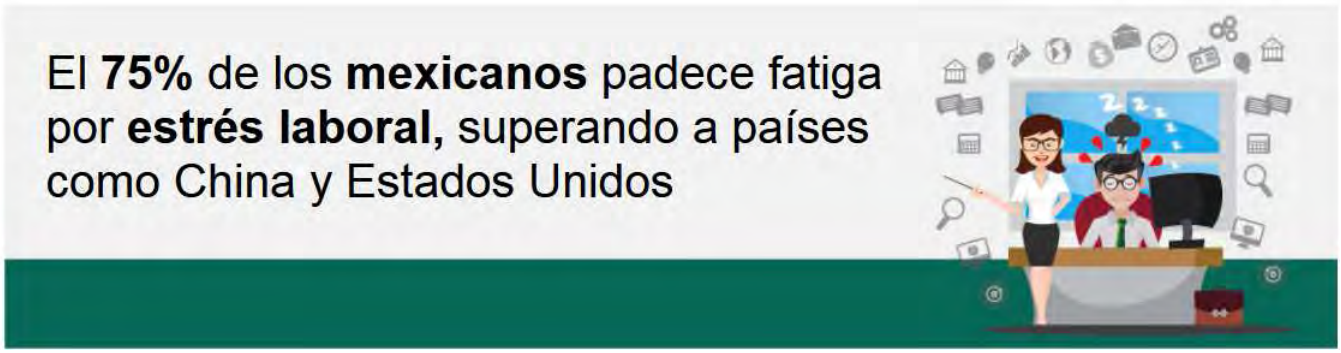


**Esthefania Ortega** es estudiante de ingeniería mecatrónica de la Universidad de Sonora, con un enfoque hacia el área de diseño, ciencia de datos y biomédica, miembro y secretaria de la sección WIE (Women in Engineering) de la IEEE rama Búhos.

**Edna Baquera** Recibió el grado de Ingeniera Industrial y de Sistemas de la Universidad de Sonora realizando sus estudios en Hermosillo, Sonora y Lyon, Francia. Actualmente se encuentra realizando el posgrado de maestría en Ingeniería, Sistemas y Tecnología en la Universidad de Sonora con especialidad en sistemas productivos y salud ocupacional.

© 2021 - Sociedad Mexicana de Inteligencia Artificial

**Jesus Pacheco** Recibió la licenciatura en Ingeniería Electrónica con un enfoque en sistemas digitales y el grado de Maestro en Ciencias de la Computación por el Instituto Tecnológico de Hermosillo y el grado de Ph. D., por el Departamento de Ingeniería Eléctrica y Cómputo de la Universidad de Arizona. Actualmente es catedrático del Departamento de Ingeniería Industrial por la Universidad de Sonora, y es Director del Centro de Cómputo Autónomo y en la Nube del NSF. Sus intereses de investigación incluyen la seguridad cibernética para el Internet de las cosas y los sistemas ciberfísicos.



Información tomada de: `http://www.imss.gob.mx/salud-en-linea/estres-laboral`.

ARTÍCULO ACEPTADO

# Identificación Automática y Robusta de Parlantes mediante Redes Neuronales Convolucionales

José Antonio Camarena Ibarrola y Karina Mariela Figueroa Mora

## Introducción

El reconocimiento automático de individuos por su voz, también conocido como reconocimiento automático de parlantes es un problema genérico que abarca dos problemas más específicos: la verificación automática de parlantes y la identificación automática de parlantes. En la verificación de parlantes, cada individuo se presenta como alguien que tiene ciertas credenciales o permisos y su identidad debe ser corroborada utilizando su voz para tal efecto. Por otro lado, la identificación automática de parlantes es un problema mas complicado dado que se debe decidir si la señal de audio que se le presenta corresponde con la voz de alguno de los individuos conocidos por el sistema, entre los cuales puede haber cientos o incluso miles. Ambos problemas se complican cuando la señal de voz está contaminada con ruido. Cuando un sistema de reconocimiento de parlantes funciona incluso si la señal de voz es ruidosa se dice que es un sistema robusto. El presente artículo presenta una técnica innovadora para resolver el problema de identificación robusta de parlantes.

Un enfoque clásico para resolver el problema de identificación de parlantes consiste en buscar en una colección de individuos conocidos al que tenga la voz más parecida a la de aquel cuya identidad se desea descubrir. Este enfoque se conoce como el del *vecino más cercano*. Dicha búsqueda se puede realizar de manera secuencial, sin embargo cuando la lista de individuos conocidos por el sistema es larga, la búsqueda secuencial resulta inviable. Es posible buscar al más parecido o *más cercano* utilizando índices de proximidad evitando un recorrido secuencial, sin embargo, los índices de proximidad exactos sufren de un problema que se conoce como *la maldición de la dimensionalidad* que consiste en que cuando la dimensión intrínseca de los datos es demasiado alta se deben hacer demasiadas comparaciones para encontrar al vecino más cercano, a veces tantas como si se hubiera realizado una búsqueda secuencial. Al extraer características de la señales de voz obtenemos vectores de dimensión alta. Además, para implementar búsquedas se requiere de una métrica que permita comparar las características de un individuo con las de otro para poder determinar que tan similares son y esa es una comparación costosa, por estas razones creemos que debemos optar por una estrategia diferente.

Las redes neuronales son clasificadores o reconocedores que no realizan búsquedas, por lo mismo no requieren de una métrica para comparar individuos. En lugar de construir índices, las redes neuronales utilizan los datos de entrenamiento para determinar los parámetros de la red neuronal. Las redes neuronales profundas no sufren la maldición de la dimensionalidad y una vez entrenadas son muy rápidas para identificar patrones.

En cuanto a la escalabilidad, es decir la capacidad de un sistema de operar adecuadamente cuando debe trabajar con muchos datos, en nuestro caso con muchos parlantes, las redes neuronales requieren de una mayor cantidad de neuronas para poder retener la información de más parlantes y en una arquitectura donde hay una neurona en la capa de salida por cada parlante, el incremento de neuronas es lineal pero las conexiones entre neuronas de la capa de salida y la capa anterior aumenta en mucho mayor grado. En general la complejidad de tiempo y de espacio depende también del número de clases (Ej. parlantes) en todos los clasificadores.

## Revisión del estado del arte

El reconocimiento de parlantes ha sido tema de investigación desde hace más de cincuenta años. En los primeros trabajos se modelaba el tracto vocal de los parlantes como una serie de tubos concatenados y el área de sección transversal y la longitud de dichos tubos al emitir un sonido específico caracterizaba al parlante. En ese esfuerzo incluso se llegaron a tomar radiografías de personas pronunciando una vocal específica [1]. En 1988, Lieberman y Blumstein publicaron estudios sobre la sensibilidad del sistema auditivo humano en donde se muestra que percibimos mucho mejor las bajas frecuencias que las altas [2]. Desde entonces los sistemas de reconocimiento de parlantes utilizan escalas logarítmicas en el dominio de la frecuencia, como la de Mel o la de Bark. En 1995 la señal de error, obtenida mediante el cálculo de la diferencia entre la señal de voz original y la sintetizada fue utilizada por Thévenaz, and P., Hügli para reconocimiento de parlantes [3], esto se debe a que la envolvente de la señal de error se asemeja a la forma del pulso glotal el cual se sabe que es diferente para cada parlante. En 1999 el tren de pulsos glotales fue utilizado también para reconocimiento de parlantes solo que, estimándolos de una mejor manera, este trabajo también fue realizado por Plumbe *et al.* [4]. En el año 2000 Besacier y Bonastre utilizaron la energía por bandas de la señal de voz en una escala logarítmica base 2 para identificación de parlantes [5]. En 2009 Yu *et al.* usaron los coeficientes cepstrales de Mel

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

que muchos consideran como el estado del arte en cuanto a caracterización de señales de voz para efectos tanto de reconocimiento de voz como de reconocimiento de individuos por su voz. Sin embargo, en 2017 la entropía espectral determinada para cada banda crítica de Bark se encontró que permitía reconocimiento aún más robusto que el que se lograba usando los coeficientes MFCC [6]. En 2019 Bunrit *et al.* extrajeron el espectrograma para cada dos segundos de la señal de voz del parlante [7].

Un espectrograma es una imagen que muestra la manera en la que la energía se distribuye en la frecuencia y la manera en que dicha energía va cambiando en el tiempo, dicha imagen se aplica a la entrada de una red neuronal convolucional para propósitos de reconocimiento de parlante texto-independiente.

**El uso de la entropía, como característica de la señal de voz, ha sido utilizado anteriormente para reconocimiento robusto, tanto del contenido de la señal de voz como del parlante.**

## Idea original, nuestra contribución

Las Redes Neuronales Convolucionales (RNC) han demostrado ser una herramienta excepcionalmente buena en lo concerniente al reconocimiento de imágenes. Para poder utilizar una RNC para reconocimiento de parlantes necesitamos convertir la señal de voz de los parlantes en imágenes, como vimos anteriormente, Bunrit *et al.* extraen espectrogramas de la señal de voz del parlante y de esa manera resuelven este problema [7].

Los espectrogramas son imágenes donde se aprecia la manera en que el espectro de magnitudes de la señal va cambiando en el tiempo, es decir son imágenes tiempo-frecuencia donde cada pixel indica la magnitud de un coeficiente de Fourier (frecuencia) para un marco específico (tiempo). Nosotros nos propusimos utilizar entropigramas en lugar de espectrogramas. Los entropigramas son también imágenes tiempo-frecuencia solo que cada pixel de un entropigrama representa la entropía de una banda en la frecuencia para un marco específico en el tiempo.

Por estudios previos y experimentos propios sabemos que caracterizarndo la señal de voz con la entropía en lugar de la magnitud o de la energía, se logra un reconocimiento más robusto, es decir, un reconocimiento con menor número de errores en particular para señales ruidosas. A continuación explicaremos la manera en la que convertimos una señal de voz en una imagen conocida como entropigrama que muestra cómo se distribuye la entropía en la frecuencia y a la vez la manera en que esta va cambiando en el tiempo.

## Convirtiendo señales de voz en imágenes

Es práctica común cuando se trabaja con una señal de audio, dividirla en segmentos de corta duración denominados marcos de tiempo, a este proceso se le conoce como enmarcado de la señal y se hace para garantizar que la señal dentro de cada uno de estos marcos de tiempo sea lo suficientemente corta para poder asumir que es estacionaria y que por tanto la transformada discreta de Fourier permitirá extraer características confiables en el dominio de la frecuencia, a estas características se les denomina *de tiempo corto* pues corresponden a una característica extraída de un segmento corto de audio. El tamaño de estos marcos de tiempo así como su traslape determinan la resolución que nuestros entropigramas tendrán en el eje horizontal, es decir en el eje del tiempo. Nosotros usamos marcos de 30 ms con traslape de 50 %, marcos de tiempo más largos no garantizarían que la señal fuera estacionaria. Una señal estacionaria es aquella cuyos componentes de frecuencia no están cambiando en el tiempo, por otro lado, marcos más cortos romperían la regla de que al menos dos ciclos completos del componente de frecuencia más baja quepan en el marco.

En cuanto al eje vertical, es decir el eje de la frecuencia, aprovechamos los estudios psico-acústicos realizados sobre el sistema auditivo humano, usando la escala de Bark, que es una escala logarítmica que agrupa en bandas críticas las frecuencias audibles por un ser humano. Existen 25 bandas críticas de Bark, cada banda crítica corresponde a una sección de la cóclea. Sin embargo, la última banda agrupa los componentes de frecuencia desde 15.5 KHz hasta 20 KHz y la voz no tiene componentes de frecuencia tan altos.

Nuestros entropigramas tienen un tamaño entonces de 24 por 32 pixeles. Las figuras 1 y 2 muestran dos entropigramas de la palabra “cero” pronunciadas por el mismo individuo (Aaron). Las figuras 3 y 4 muestran dos entropigramas de la misma palabra pero ahora pronunciadas por otra persona (Erendira). Observe que los entropigramas de las figuras 1 y 2 son muy parecidos. También los entropigramas de las figuras 3 y 4 se parecen mucho entre ellos. Por otra parte, los dos entropigramas producidos a partir de la voz de Aaron se ven diferentes a los dos producidos a partir de la voz de Eréndira. Podemos entonces formular la hipótesis de que hay una distancia entre clases grande y una distancia intraclase pequeña y eso es lo que necesitamos para que un sistema de identificación de parlantes funcione adecuadamente.

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

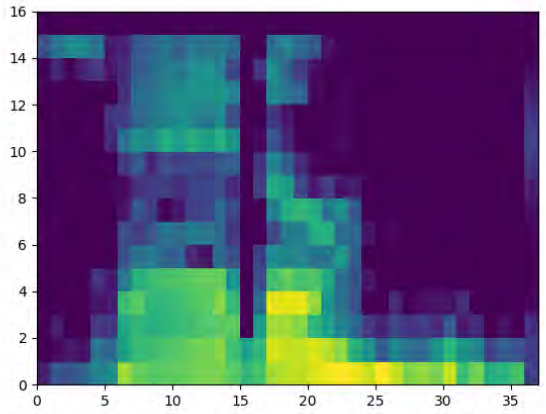



**Figura 1.** Primera elocución de Aaron.

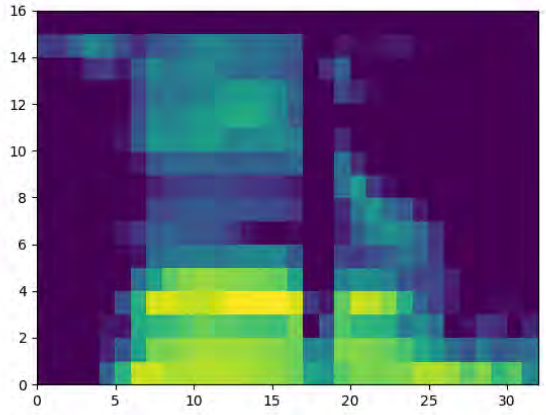



**Figura 2.** Segunda elocución de Aaron.

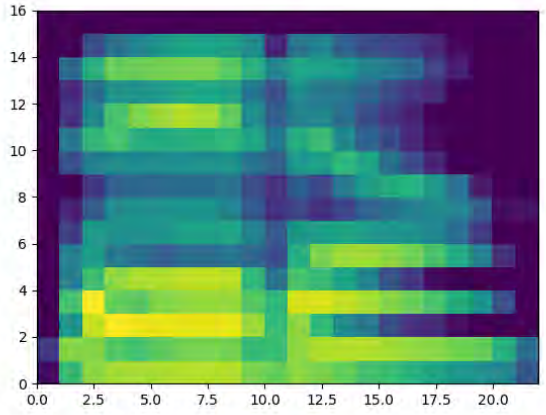



**Figura 3.** Primera elocución de Erendira.

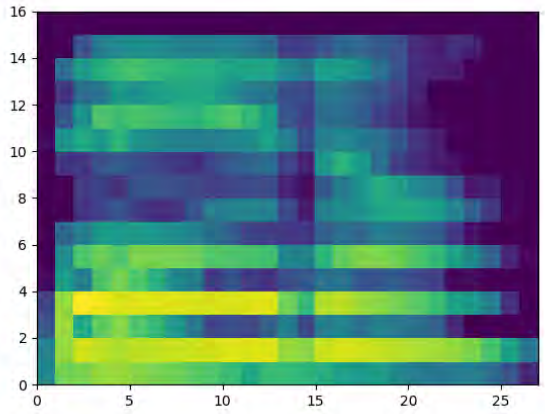



**Figura 4.** Segunda elocución de Erendira.

## Redes Neuronales Convolucionales

La convolución es la operación básica para procesamiento de imágenes. Se pueden diseñar filtros para suavizar una imagen, para aumentar el contraste o para enfatizar los bordes, por poner algunos ejemplos. El efecto que tendrá un filtro sobre una imagen depende de los coeficientes del filtro. Una ventana de un tamaño predefinido se mueve por toda la imagen y para cada posición de la ventana, el pixel que queda justo en el centro de la ventana es modificado de modo que su nuevo valor depende del resultado de la convolución de los valores de los pixeles que quedan dentro de la ventana con los coeficientes del filtro. Las redes neuronales convolucionales tienen capas que hacen lo mismo que los filtros, solo que los coeficientes o *parámetros* para hacer la convolución no tienen que conocerse a priori sino que son *aprendidos* durante el *entrenamiento* mediante un algoritmo de optimización como el de *propagación hacia atrás*. Para cada posible posición de la ventana hay una neurona en la capa convolucional, la neurona corresponde al pixel que queda en el centro de la ventana y todos los pixeles dentro de la ventana son entradas de dicha neurona. Existen pues, muchas conexiones entre la capa de entrada y la primer capa convolucional y uno pensaría que entonces hay una cantidad enorme de parámetros que determinar durante el entrenamiento, sin embargo, en realidad los parámetros son los mismos para todas las ventanas pues son los coeficientes del filtro al que nos referimos, por tanto el número de parámetros o pesos entre la capa de entrada y la primer capa convolucional no depende del tamaño de la imagen sino del tamaño de la ventana. Los parámetros aprendidos durante entrenamiento hacen que la primer capa convolucional funcione como un filtro que descubre formas primitivas en la imagen, sin embargo, hay diferentes formas primitivas en una imagen, la solución es agregar capas convolucionales (es decir filtros) paralelas, cada filtro paralelo busca una forma primitiva

diferente. Normalmente la capa de neuronas que sigue de una capa convolucional es una capa que reduce el tamaño de la imagen resultante para reducir el esfuerzo computacional que implican las capas convolucionales posteriores, las cuales sirven para encontrar formas más complejas puesto que se conforman con las formas más simples descubiertas por las capas convolucionales que les preceden.

## Nuestra Red Neuronal Convolucional

Las imágenes tienen un tamaño de $24 \times 32$ pixeles y éstas son aplicadas a la capa de entrada de la red neuronal, que es una capa convolucional con un tamaño de ventana $3 \times 3$ y con 16 filtros paralelos. Como hay $3 \times 3$ pixeles en la ventana y cada pixel tiene 4 componentes (Rojo, verde, azul y transparencia), entonces cada neurona de la primera capa convolucional tiene $3 \times 3 \times 4 = 36$ entradas más una adicional de sesgo y como son 16 filtros paralelos, entonces la primera capa cuenta con $37 \times 16 = 592$ parámetros. La segunda capa sirve para reducir el tamaño de la imagen antes de pasarla a la siguiente capa, se trata de una capa del tipo *maxPool* con ventanas de $2 \times 2$ de modo que las imágenes se reducen a un tamaño de $11 \times 15$ pixeles, las cuales son procesadas por la siguiente capa que es convolucional y tiene 32 filtros paralelos, de modo que cada neurona de esa capa tiene $3 \times 3 \times 16 = 144$ entradas más la de sesgo, por lo que dicha capa tiene $145 \times 32 = 4640$ parámetros. La siguiente es otra capa del tipo *maxPool* con ventanas de $2 \times 2$ de modo que las imágenes se reducen a un tamaño de solo $4 \times 6$. La siguiente capa simplemente convierte los tensores 3D a tensores 1D. La siguiente capa es densa, es decir conecta las $4 \times 6 \times 32 = 768$ neuronas del tensor 1D con cada una de sus 64 neuronas, cada una de las cuales tiene 768 entradas más la de sesgo, lo cual implica $769 \times 64 = 49,216$ parámetros adicionales. Finalmente, la capa de salida también es densa y conecta las 64 neuronas de la capa anterior con sus 21 neuronas, cada una de las cuales tiene 64 entradas más la de sesgo, lo que agrega $65 \times 21 = 1,365$ parámetros a la red. La capa de salida utiliza la función de activación *Softmax* dado que usamos para nuestros experimentos 21 parlantes. La función Softmax se encarga de que la salida de cada neurona tenga un valor entre 0 a 1 que es la probabilidad de que el entropigrama en la capa de entrada corresponda con un parlante específico. En total son 55,813 parámetros que se deben optimizar durante el entrenamiento de la red neuronal, este número de parámetros es relativamente bajo para una red profunda, gracias al hecho de que nuestros entropigramas son imágenes muy pequeñas.

## Experimentos

Para evaluar el funcionamiento de nuestro identificador de parlantes utilizamos una colección de grabaciones de la voz de 21 personas de nacionalidad mexicana que incluía hombres y mujeres con edades entre los 18 y los 30 años. Cada individuo pronunció 34 palabras en español en 4 ocasiones cada palabra, de modo que la colección está formada por $21 \times 34 \times 4 = 2,856$ grabaciones, la colección se puede descargar [1]. Utilizamos las primeras 3 elocuciones de cada palabra como conjunto de entrenamiento y las cuartas elocuciones formaron el conjunto de prueba. Entonces se obtuvieron $21 \times 34 \times 3 = 2,142$ entropigramas para entrenar la red y $21 \times 34 = 714$ entropigramas para realizar pruebas.

Implementamos la red neuronal convolucional en Python utilizando la librería Keras, la cual está construida sobre TensorFlow. Usamos el ambiente de desarrollo Google Colab. Para el entrenamiento de la red, es decir para la determinación de los parámetros óptimos, iniciamos con valores aleatorios y usamos el optimizador *adam*, el cual es un algoritmo de gradiente descendente. Obtuvimos una precisión de 91 % con el conjunto de entrenamiento y de 80 % con el conjunto de prueba. Para comparar el desempeño de nuestro identificador de parlantes que usa entropigramas, con uno equivalente que usara espectrogramas, repetimos todo el proceso de extracción de las imágenes a partir de los audios, usando la misma resolución que para los entropigramas solo que ahora cada pixel representa cantidad de energía contenida en una banda crítica de Bark y marco específicos, en lugar de cantidad de información (entropía); re-entrenamos la red neuronal y obtuvimos una precisión de 82 % con el conjunto de entrenamiento pero solo de 63 % con el conjunto de prueba.

## El uso de redes neuronales convolucionales, en combinación con entropigramas, mejoran el reconocimiento robusto de parlantes

Hasta ahora, lo que hemos corroborado es que los entropigramas sí pueden ser utilizadas como imágenes para identificar individuos por su voz, e incluso funcionan mejor que los espectrogramas cuando el tamaño de las imágenes es tan pequeño como 24 por 32 pixeles. Sin embargo, nuestra principal hipótesis era que los entropigramas serían mejores que los espectrogramas para señales ruidosas, es decir que permitirían identificación

[1] *dep.fie.umich.mx/∼camarena/dsp/elocuciones21.tar.gz*

*robusta* de parlantes. Para corroborarlo, mezclamos todos los audios, tanto de prueba como de entrenamiento, con ruido blanco hasta obtener una relación señal a ruido de tan solo cero decibeles. Esto significa que la energía de la señal de voz y la energía del ruido es igual. Al escuchar esos audios ruidosos, incluso cuesta trabajo reconocer la voz en ellos. Entonces, repetimos de nuevo todo el proceso de determinación de entropigramas, de espectrogramas y de entrenamiento de la red neuronal, y los resultados que obtuvimos con entropigramas fue de una precisión del 90 % con el conjunto de entrenamiento y de 79 % con el conjunto de prueba, en cambio, cuando usamos espectrogramas, la precisión con el conjunto de entrenamiento fue de solo 62 % y con el conjunto de prueba de apenas 55 %. Cabe mencionar que dado que antes del entrenamiento de la red neuronal todo los parámetros tenían valores aleatorios, los experimentos se repitieron cinco veces y los resultados mencionados son en realidad promedios. Las varianzas son todas inferiores a 0.5, por ejemplo, al experimentar con entropigramas de señales de voz ruidosas, la precisión mínima obtenida fue de 77 % y la máxima fue de 81 % la media reportada es 79 %.

## Agradecimientos

Los autores agradecen a la Coordinación de la Investigación Científica de la Universidad Michoacana de San Nicolás de Hidalgo por el apoyo otorgado para el desarrollo de proyectos relacionados al tema aquí presentado, así como a los revisores de la revista Komputer Sapiens por sus valiosos comentarios para la redacción del presente artículo de divulgación científica.

## Conclusiones

Planteamos la hipótesis de que los entropigramas servirían como imágenes, que al ser aplicadas a una red neuronal convolucional, nos permitirían identificar al parlante de cuya voz se extrajo dicho entropigrama. Nuestro experimentos confirmaron esta hipótesis, incluso mejoraron el desempeño de los espectrogramas para este propósito para imágenes de baja resolución de tan solo 24 por 32 pixeles. También planteamos la hipótesis de que los entropigramas permitirían implementar sistemas de identificación de parlantes robustos al ruido. De nuevo nuestros experimentos confirmaron la hipótesis, pues el identificador basado en entropigramas sufrió poca afectación al trabajar con señales ruidosas respecto a su funcionamiento con señales libres de ruido, puesto que su precisión bajó de 80 % a 79 %, en contraste, el identificador de parlantes basado en espectrogramas fue severamente afectado por el ruido, ya que bajó su precisión de 63 % a tan solo 55 %.

Es claro que el identificador basado en redes neuronales convolucionales sufre de sobreentrenamiento. Para lidiar con ello agregamos una capa *Dropout*, entre las dos capas densas. Dropout es en realidad una técnica que obliga al algoritmo de entrenamiento a olvidar los valores de los parámetros de la capa anterior durante el entrenamiento cada número de iteraciones con cierta probabilidad. La probabilidad que usamos fue de 0.2 puesto que fue el valor que pareció funcionar mejor. Sin embargo existen otras técnicas para lidiar con el sobreentrenamiento. Una que valdría la pena probar, se conoce como *Aumento de datos* y consiste en realizar modificaciones aleatorias en las imágenes, como: traslaciones, rotaciones escalamientos, cambio en el brillo o contraste, etc. ✷

## REFERENCIAS

1. Rabiner, L.R., Schafer, R.W. (1978). Digital processing of speech signals. Prentice Hall.

2. Lieberman, P., Blumstein, S.E. (1988). Speech physiology, speech perception, and acoustic phonetics. Cambridge University Press.

3. Thévenaz, P., Hügli, H. (1995). Usefulness of the lpc-residue in text-independent speaker verification. *S*peech Communication **17**(1-2), 145–157.

4. Plumpe, M.D., Quatieri, T.F., Reynolds, D.A. (1999). Modeling of the glottal flow derivative waveform with application to speaker identification. *I*EEE Transactions on Speech and Audio Processing, 7(5), 569–586.

5. Besacier, L., Bonastre, J.F. (2000). Subband architecture for automatic speaker recognition. *S*ignal Processing, 80(7), 1245–1259.

6. Camarena-Ibarrola, A., Luque, F., Chavez, E. (Nov 2017). Speaker identification through spectral entropy analysis. En: *2*017 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), 1–6. https://doi.org/10.1109/ROPEC.2017.8261607.

7. Bunrit, S., Inkian, T., Kerdprasop, N., Kerdprasop, K.: Text-independent speaker identification using deep learning model of convolution neural network. *I*nternational Journal of Machine Learning and Computing, 9, 143–148 (04 2019). https://doi.org/10.18178/ijmlc.2019.9.2.778.

## SOBRE LOS AUTORES

**José Antonio Camarena Ibarrola** es doctor por la Universidad Michoacana de San Nicolás de Hidalgo, estudió la maestría en Ciencias Computacionales en el Instituto Tecnológico de Toluca y sus intereses de investigación están en el área de reconocimiento de patrones, procesamiento de señales y procesos estocásticos.

**Karina Mariela Figueroa Mora** obtuvo el grado de doctora en Ciencias de la Computación en la Unversidad de Chile, estudió la maestría en Ingeniería Eléctrica en la Universidad Michoacana de San Nicolás de Hidalgo, sus intereses de investigación están en el area de búsquedas de proximidad y recuperación de información.

*Publique en Komputer Sapiens*

ARTÍCULO ACEPTADO

# Sistemas y algoritmos de Inteligencia Artificial para combatir la violencia de género

Obed Campos Solano, Pablo Pancardo y José Adán Hernández Nolasco

## Introducción

Los dispositivos móviles son plataformas que pueden ser muy útiles para instalar aplicaciones que combaten la violencia de género. Gobiernos, organismos y empresas de todo el mundo han desarrollado este tipo de aplicaciones especialmente en defensa de las mujeres. En México, la violencia de género afecta al 60 % de las mujeres e impacta en el bienestar social, físico, psicológico y económico de éstas. Es por ello que existen aplicaciones de seguridad personal para solicitar ayuda en una situación de emergencia.

En India; "848 mujeres son acosadas, violadas y asesinadas cada día" [1] y por ello, comprobamos que la violencia es una situación grave que pone en peligro la vida de muchas mujeres. La violencia de género no es un problema local, está presente en las ciudades primermundistas.

Aunque los datos estadísticos muestran que la mayoría de los incidentes son contra las mujeres, existen estudios que revelan que la violencia puede presentarse incluso contra hombres. En un experimento social creado por la Universidad del Sur de California [2], los participantes debían llevar dispositivos para medir su ritmo cardíaco y niveles de estrés. El experimento mostraría si los conflictos de la vida diaria afectaban la relación entre las parejas al llegar a casa. Los resultados mostraron que ambos se veían afectados emocionalmente al final del día.

En el estado del arte, encontramos tecnologías que se han utilizado para detectar violencia, con resultados prometedores. Estas tecnologías incluyen Algoritmos de inteligencia artificial que varios autores han empleado para analizar sonidos y determinar si el audio contiene características importantes (como gritos, disparos, etc.). Otros autores utilizaron estas características para analizar películas y crear una base de datos con el fin de evitar que los niños vean contenido violento.

Después de examinar el estado del arte existente sobre la violencia de género, hicimos una selección para clasificar los artículos según la rama a la que pertenecen. Cabe mencionar que hay muchos estudios en la India. En Nueva Delhi ocurrió un incidente en 2012: una joven fue violada por 7 hombres en un camión de pasajeros, muriendo dos semanas después.

La noticia se difundió internacionalmente. Debido a esto se realizó una encuesta para medir la efectividad del botón de pánico, un requisito que actualmente es ley en el país. Ésta establece que todos los smartphones que se vendan, deben tener un botón de pánico para proteger a las mujeres.

Nuestro principal objetivo es analizar los sistemas, tecnologías y métodos que se han utilizado para detectar la violencia de género en tiempo real. Hasta el momento, no existe ninguna revisión sobre el uso de la tecnología para detectar violencia física y verbal. Por ello, clasificamos este trabajo en dos apartados:

1. Sistemas y Aplicaciones para detectar la violencia de género.
2. Algoritmos de Inteligencia Artificial para detectar la violencia.

## Clasificación del Estudio

### Sistemas y aplicaciones para detectar la violencia de género

Tras el análisis, hemos encontrado trabajos con propuestas para detectar la violencia de género. Fueron realizados con el fin de ayudar a las mujeres en situaciones de peligro: iMace: Esta aplicación emite una fuerte alarma cuando el teléfono es agitado vigorosamente, para así ahuyentar al atacante. Además, se toman fotos automáticamente para enviarlas como prueba a la policía, junto con la localización de la víctima. SMARISA es un anillo inteligente. Los autores pretenden utilizar el Internet de las Cosas (IoT) como una solución novedosa para la seguridad de las mujeres. Cuando se activa, el teléfono envía una imagen a los contactos de emergencia.

[1] propone el diseño de una pulsera GPS. Se conecta vía Bluetooth con un smartphone. Al detectarse valores elevados de temperatura corporal o frecuencia cardíaca, se envía la ubicación a la policía y contactos de emergencia a través de un SMS.

Glenson Toney ha diseñado e implementado un brazalete. Mediante un movimiento de mano, Google Maps envía la ubicación cada 30 segundos y un video en vivo a la policía.

FEMME es un dispositivo diseñado especialmente para mantener la seguridad de las mujeres y ayudarlas en situaciones de emergencia. El dispositivo detecta si hay una cámara cerca que podría utilizarse para acosar a la víctima: "este dispositivo es un receptor de radiofrecuencias que analiza las señales electromagnéticas en busca de aquellos dispositivos electrónicos que suelen emitir ese tipo de señales, como las cámaras espía.".

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

**Aunque los datos estadísticos muestran que la mayoría de los incidentes son contra las mujeres, existen estudios que revelan que la violencia puede presentarse incluso contra hombres.**

Naeemul Islam implementó un sistema que envía la ubicación GPS a los contactos de emergencia cuando se presiona un botón.

En su tesis doctoral, Parth Sethi propone evitar la agresión sexual mediante el uso de un dispositivo en forma de parche conectado a un teléfono vía Bluetooth, que lee movimientos inusuales en la ropa, como los movimientos de fuerza de un agresor. El sistema se enciende y el dispositivo emite un sonido o un olor penetrante para alejarlo.

Google Maps es un sistema muy práctico para utilizar la localización en tiempo real. Se ha diseñado un sistema inteligente para proteger a las mujeres que han sufrido algún tipo de violencia por parte de un agresor. Para que éste funcione, el agresor tuvo que haber atacado a la víctima, y es necesaria la intervención policial para poner una tobillera en ambas personas. Esto para monitorear la ubicación en tiempo real. Se rastrean ambas señales y es posible detectar cuando se sobrepasa la distancia establecida. Cuando esto ocurre, se notifica a la víctima y a los agentes de policía.

WoSApp es una aplicación para android que tiene como función principal informar a los contactos de emergencia. Esta aplicación puede activarse al agitar el smartphone; dicho de forma más sencilla, es un botón de pánico digital.

Safe Sole (Figura 1) es una suela de zapato Bluetooth que enciende con solo dar tres golpecitos con el talón enviando la ubicación GPS a los contactos de emergencia. Para medir los golpecitos, el sistema contiene un sensor de presión que califica la fuerza aplicada al golpecito.

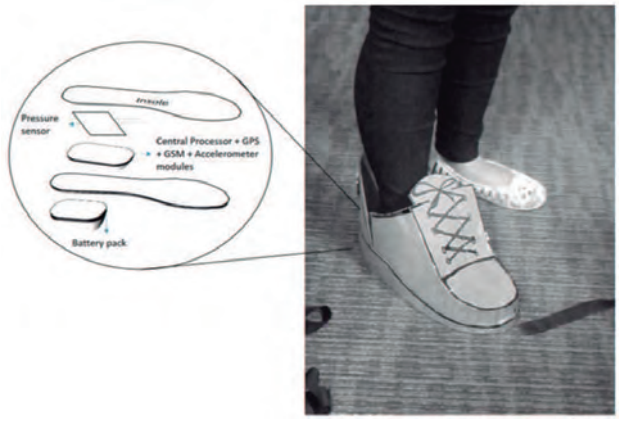



**Figura 1.** Prototipo Experimental *Safe Sole*.

Entre todos los algoritmos de inteligencia artificial encontrados en los trabajos para detectar la violencia, ninguno se utilizó específicamente en violencia de género. Pero cabe mencionar que la mayoría de los autores realizaron sus investigaciones hace tiempo. Sin embargo, los trabajos más recientes incluyen estos mismos algoritmos. Esto significa que algunos de estos trabajos no han perdido su precisión con el paso del tiempo y siguen dando resultados sobresalientes. Los trabajos seleccionados contienen una buena proporción de temas para detectar la violencia de género.

Hemos ordenado los datos de forma descendente, teniendo en cuenta los métodos más utilizados, y hemos analizado en detalle los eventos exactos que detecta cada algoritmo, así como el conjunto de datos que utiliza cada proyecto y el nivel de precisión en la detección reportado por cada autor para proporcionar un trabajo más completo.

Observamos que la mayoría de los autores hacen uso del algoritmo de clasificación máquina de soporte vectorial (SVM), que puede utilizarse para extraer características de audio, como la entropía de energía, la energía de tiempo corto, la tasa de cruce por cero, el flujo espectral y el espectro de la señal [3]. Sin embargo, aunque el SVM es uno de los algoritmos más utilizados para clasificar, hay muchas otras técnicas utilizadas para el mismo propósito, como el Modelo de Mezcla Gausiana (GMM) [4] o el Coeficiente Cepstral de Frecuencia Mel (MFCC) [5].

Además, observamos que la forma más eficiente de obtener resultados precisos es combinando las fuentes de audio y vídeo (Figura 2) [6]. El inconveniente que vemos con este tipo de combinación es que siempre que se produce una situación de violencia de género, la víctima puede no tener el tiempo o la posibilidad de sacar el teléfono y empezar a grabar video. Sin embargo, funciona perfectamente con los sistemas de Circuito Cerrado de Televisión (CCTV), dado que las cámaras están instaladas en un lugar inamovible de la habitación o edificio (incluso con CCTV instalados dentro de unidades de transporte público) y están grabando todo el tiempo.

**Figura 2.** Sistema propuesto por [6].

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

Otros estudios se han hecho para detectar peleas, armas como cuchillos o armas de fuego [6, 7], clasificar escenas violentas de clips de películas y lo más importante, detectar emociones negativas, y violencia en el discurso para prevenir la violencia doméstica. Todos los autores proponen el uso de algoritmos de inteligencia artificial y los aplican para detectar o prevenir una situación de riesgo simplemente analizando las señales de audio del discurso [8, 9] y otros, utilizaron estos algoritmos para clasificar emociones negativas como la ira.

Aunque estos trabajos de investigación tienen excelentes propuestas, nuestro análisis sugiere que debería haber un método más preciso para detectar estas situaciones delicadas y peligrosas; una combinación de algoritmos o técnicas basadas en la inteligencia artificial para aumentar el porcentaje de detección/clasificación.

***Aplicaciones de Seguridad Personal en México***

Las aplicaciones de seguridad personal contra la violencia de género deben ser evaluadas en términos de usabilidad. Sin embargo, aunque existen métodos para evaluar la usabilidad y eficacia de aplicaciones móviles, las particularidades de las aplicaciones de seguridad personal referentes al escenario y contexto en el que se utilizan, hacen que las métricas de calidad no sean suficientes. Y es que las aplicaciones de violencia de género requieren una evaluación diferente para determinar su eficacia y usabilidad. Hasta donde sabemos, nadie ha evaluado la eficacia de estas aplicaciones en entornos reales, sólo su interacción. Antes de analizar la usabilidad, queremos describir algunas aplicaciones desarrolladas para mujeres mexicanas.

**Alerta Rosa:** Se conecta por Bluetooth con la pulsera Mujer Segura Alerta Rosa y ofrece un botón de pánico que envía una señal al C5 y permite identificar geográficamente la alerta, para que los servicios de seguridad atiendan la llamada. Estas pulseras son distribuidas por las instancias locales de la Ciudad de México. Si no se tiene, o se descarga la batería, el sistema es inútil. Como punto a favor, sólo pulsando el botón, se envían las alertas. Es vital tenerla funcionando en segundo plano.

**Mujer Alerta:** Desarrollada por el Gobierno de Veracruz, dirigida a mujeres veracruzanas, para brindarles asesoría y asistencia, con base en la Declaratoria de Alerta de Violencia de Género actualmente establecida, cuyos propósitos son: 1) Identificar tipos y modalidades de violencia contra las mujeres. 2) Ubicar lugares a los que puedes acudir para recibir asesoría y denunciar estos actos en tu municipio. 3) Botón de pánico cuando te encuentres en una situación de emergencia. Esta aplicación requiere conexión a datos móviles y geolocalización, lo que puede ser un inconveniente cuando no se tiene acceso a internet.

**Mujeres Seguras:** Desarrollada por el gobierno de Sonora. Este sistema tiene dos opciones: 1. Añadir contactos de confianza. 2. Botón de pánico. Una desventaja es que no existe la opción de notificar a los contactos de emergencia. Otra es que el sistema no se puede activar a través de gestos. Aunque el sistema carece de funcionalidad, se comunica inmediatamente con las autoridades pulsando el botón, lo cual es una función útil.

**No Estoy Sola:** Desarrollada por el gobierno de Ciudad Juárez, su única función es presionar el botón de pánico para enviar una alerta a los contactos de emergencia.

La desventaja es que hay que presionarlo durante 3 segundos para que se envíe la alerta. Esto no es conveniente ya que las víctimas pueden no tener esos tres segundos.

**SOSMEX:** Su objetivo principal es alertar a los contactos de emergencia. Cuando se pulsa el botón, la aplicación envía un SMS que contiene la ubicación actual. El sistema tiene la opción de llamar a los contactos. Cuando se instala, se crea un acceso directo en la barra de notificaciones para garantizar un acceso rápido. Se puede activar pulsando cinco veces el botón de encendido o agitando el teléfono las mismas cinco veces. Para iniciar el sistema, es necesario desbloquear el smartphone, abrir la aplicación y pulsar el botón; o bien, pulsar el acceso directo en la barra de notificaciones. Esto es un punto débil dado el número de pasos a seguir; como ventaja, la infraestructura del sistema está compuesta por servidores que manejan las llamadas (se envía un comando al servidor para hacer la llamada. El dispositivo no hace las llamadas).

**vgenero:** Desarrollada HEXACOM (Argentina). El sistema envía una alerta a los contactos de emergencia, así como a las autoridades municipales de la ciudad. Para el registro, los datos ingresados son validados por las entidades correspondientes; de lo contrario, no es posible. Esto es una desventaja, ya que en el momento en que se necesite un registro rápido, no habrá esta posibilidad.

***Aplicación vs Escenario***

Existen estudios que analizan la usabilidad de las aplicaciones móviles. Es decir, determinan lo fácil o difícil que es interactuar con la aplicación, de manera que cumpla con el propósito para el que fue desarrollada, dependiendo del escenario y las condiciones para una situación particular.

Según [10], las dimensiones a tener en cuenta son: satisfacción del usuario, eficiencia y eficacia. La primera, se refiere a lo satisfecho que está el usuario haciendo uso de la aplicación, y como ésta suple la necesidad para la que ha sido creada. La segunda, se refiere a qué tan eficiente es la aplicación de acuerdo a su uso. Por ejemplo, WhatsApp cumple con este propósito porque todas las opciones funcionan correctamente. La tercera, es la eficacia de las funciones. Volviendo al ejemplo anterior, la aplicación cumple este objetivo porque no ocupa

demasiado almacenamiento y los mensajes llegan instantáneamente.

[10] hace una evaluación de las aplicaciones en escenarios peligrosos. Se experimentó con 30 personas que instalaron una aplicación de seguridad en sus teléfonos. Se les entrevistó antes (preApp) y después (postApp) de instalar dichas aplicaciones. Uno de los resultados del estudio preApp fue que algunos usuarios no confiaban en la eficacia de las aplicaciones y creían que no podrían utilizarlas en escenarios reales.

## Trabajos Futuros

Este trabajo es un avance de una revisión mayor. Se ha realizado para mostrar el análisis que tenemos hasta ahora. En un futuro trabajo se publicará un novedoso método de clasificación que pueda detectar la violencia entiempo real. Nuestro objetivo es desarrollar un método que pueda identificar la presencia de voces de hombre y mujer en audio. El objetivo específico que buscamos es separar la violencia en general de la violencia de género porque estamos tratando de limitar el alcance de nuestra investigación a un área específica.

El método que tratamos de construir es una combinación de dos algoritmos de inteligencia artificial, con el propósito de determinar en tiempo real la violencia de género. Este tipo de violencia se da sólo cuando una mujer ejerce violencia contra un hombre o viceversa. Pretendemos utilizar el reconocimiento de voz para analizar las palabras, y compararlas con la semántica dentro del ámbito de la violencia de género (malas palabras, gritos, etcétera).

A corto plazo, pretendemos utilizar inteligencia artificial ya que ha demostrado obtener resultados eficientes y precisos con estas situaciones. Con el uso de estas tecnologías y métodos, las soluciones se pueden medir en tiempo real, y no hay necesidad de intervención humana. De este modo, se elimina el uso de un botón digital, movimientos para activar el sistema, y el riesgo de que, si la víctima no tiene tiempo para hacer esos movimientos, el sistema no se active.

Los métodos de detección actuales presentan una carencia de características importantes que harían más efectiva esta tarea. Estas características son métodos capaces de actuar de forma automática, sin necesidad de ser activados por algún disparador, simplemente por inteligencia artificial. La solución debe ser en tiempo real, sin aplicar la minería de datos para analizar la situación, pretendemos utilizar inteligencia artificial para detectar el ataque.

## Discusión y Conclusiones

Después de haber realizado el análisis de la tecnología existente en la actualidad para detectar violencia de género, podemos concluir que la intención de los autores en desarrollar los sistemas y aplicaciones es muy buena, ya que atacan una problemática social existente en todos los países del mundo, pero se observa que la practicidad para operar dichos métodos se encuentra todavía en fases de desarrollo.

Por otro lado, el uso de inteligencia artificial para automatizar y clasificar eventos es de mucha ayuda, ya que los algoritmos pueden clasificar características que, mediante un proceso convencional no sería posible. Aunque, de igual forma que los métodos anteriores, todavía no hay mucho desarrollo en el área, a excepción de pocos que han logrado más del 90

A pesar del gran crecimiento que se ha ido logrando en otros países en materia de seguridad y prevención de ataques en contra de la mujer, en México solo se han desarrollado aplicaciones móviles de parte de algunos gobiernos municipales y la mayoría de ellos forman parte de un sistema para denuncias ciudadanas, por lo que la parte de la seguridad hacia la mujer es solamente un elemento más del sistema.

De las aplicaciones dedicadas a proteger a las mujeres, ninguna de ellas integra inteligencia artificial, por lo que es necesario que la persona acceda a la aplicación para activar el botón de pánico virtual, Lo que conlleva a desbloquear el teléfono en primer lugar, perdiendo valiosos segundos que en un momento así, son vitales.

Esperamos en un corto plazo terminar el desarrollo de una tecnología que será de mucho apoyo y trabaje en coordinación con las autoridades locales y sistemas de videovigilancia o CCTV para detectar ataques en tiempo real.✳

## REFERENCIAS


1. Harikiran, G., Menasinkai, K. y Shirol, S. (2016). Smart security solution for women based on Internet Of Things (IOT). En *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*. Chennai, India. 3551-3554.

2. Timmons, A., Chaspari, T., Han, S., Perrone, L., Narayanan, S. y Margolin, G. (2017). Using multimodal wearable technology to detect conflict among couples. *Computer*, 50(3), 50-59.

3. Giannakopoulos, T., Kosmopoulos, D., Aristidou, A. y Theodoridis, S. (2006). Violence content classification using audio features. En *Hellenic Conference on Artificial Intelligence*. Berlin, Heidelberg, 502-507.

4. Lee, Y., Kim, K., Han, D. y Ko, H. (2012). Acoustic and visual signal based violence detection system for indoor security application. En *IEEE International Conference on Consumer Electronics (ICCE)*. Las Vegas Nv, USA. 737-738.

5. Bautista-Duran, M., García-Gomez, J., Gil-Pita, R., Sánchez-Hevia, H., Mohino-Herranz, I.y Rosa-Zurera, M. (2017). Acoustic Detection of Violence in Real and Fictional Environments. En *ICPRAM*. 456-462.

6. Andersson, M., Ntalampiras, S, Ganchev, T., Rydell, J., Ahlberg, J., Division of Information Systems, FOI Swedish Defence Research Agency, Sweden, L. y Fakotakis, N. (2010). Fusion of acoustic and optical sensor data for automatic fight detection in urban environments. En *13th International Conference on Information Fusion*. Edinburg, UK. 1-8.

7. Grega, M., Matiolański, A., Guzik, P. y Leszczuk, M. (2016). Automated detection of firearms and knives in a CCTV image. *Sensors*, 16(1), 47.

8. Shulby, C., Pombal, L., Jordao, V., Ziolle, G., Martho, B., Postal, A. y Prochnow, T. (2018). Proactive Security: Embedded AI Solution for Violent and Abusive Speech Recognition. En *7th Brazilian Conference on Intelligent Systems (BRACIS)*. Sao Paulo, Brazil. 248-253.

9. Subramani, S., Vu, S. y Wang, H. (2017). Intent classification using feature sets for domestic violence discourse on social media. En *4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*. Mana Island, Fiji. 129-136.

10. Just, M., Chandok, H., Sampangi, R., Hawkey, K., Willis, A., Samuthiravelu, J., Gill, D. y Altair, M. (2019). Personal safety app effectiveness. En *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow Scotland, UK. 1-6.

## SOBRE LOS AUTORES

**Obed Campos Solano** es un estudiante de Doctorado en Ciencias de la Computación en la División Académica de Ciencias y Tecnologías de la Información en la Universidad Juárez Autónoma de Tabasco (UJAT). Obtuvo su grado de maestría en Tecnologías de Internet en la Universidad Autónoma de Guadalajara en 2010. Sus intereses de investigación incluyen procesamiento de lenguaje natural, redes neuronales y lógica difusa.

**Pablo Pancardo** es Doctor en Ciencias de la Computación por la Universidad Juárez Autónoma de Tabasco (UJAT). Actualmente es Profesor-Investigador y responsable del Cuerpo Académico "Sensado y Procesamiento Inteligente de Datos" en la UJAT. Cuenta con más de 40 publicaciones en conferencias internacionales y revistas de investigación relacionadas con el cómputo ubicuo para el cuidado de la salud y el bienestar. Ha participado en varios proyectos de investigación que involucran la adopción del cómputo móvil ubicuo y redes de sensores. Sus intereses de investigación incluyen la inteligencia ambiental, la lógica difusa, la Internet de las cosas y la Interacción humano-computadora.

**José Adán Hernández Nolasco** Ingeniero en Electrónica y Comunicaciones por la Universidad Autónoma de Nuevo León, M. en C. en Ingeniería Electrónica [Telecomunicaciones] por el Instituto Tecnológico y de Estudios Superiores de Monterrey, Doctor en Ciencias con especialidad en Óptica por el Instituto Nacional de Astrofísica Óptica y Electrónica. Miembro del Sistema Nacional de Investigadores, con 23 años como Profesor Investigador en la Universidad Juárez Autónoma de Tabasco. Ha sido autor o coautor en 25 publicaciones, con alrededor de 30 participaciones en congresos. Sus intereses principales en investigación son: Inteligencia ambiental, lógica difusa, redes neuronales y óptica.

*Sis Bot es un bot de chat que puede proporcionar servicios de información 24 horas al día, 7 días a la semana para sobrevivientes de violencia, accesible a través de su dispositivo móvil o una computadora.*

ARTÍCULO ACEPTADO

# Aprendizaje de máquina para la identificación del comportamiento de perros de búsqueda y rescate

Rodrigo Chi Ozalde, Humberto Pérez Espinosa y Jorge Ríos Martinez

## Introducción

Cuando ocurre un desastre natural frecuentemente es necesario encontrar a personas que se encuentran desaparecidas que pueden estar en una situación de riesgo. En estos casos el tiempo es un factor clave que puede ser la diferencia entre la vida y la muerte para estas personas. Para reducir los tiempos de búsqueda y localización en estos eventos, los servicios de emergencia, policía y otros grupos de rescate implementan binomios caninos.

En México 11 de los 32 estados que lo conforman se encuentran en zonas sísmicas de mediano y alto riesgo [1]. El área con mayor riesgo se concentra en la costa occidental del país a lo largo de los bordes de varias placas tectónicas.

Se estima que ocurre un sismo de magnitud 7.5 grados o mayor en la escala de Richter y cinco sismos de magnitud 6.5 grados o mayor cada cuatro años.

Para localizar a las personas desaparecidas en estos desastres se emplean equipos formados por un perro de búsqueda y rescate (PByR) y un entrenador experto en darle instrucciones al perro para realizar tareas.

Estos equipos o binomios son altamente efectivos. En eventos recientes (sismo México 2017 7.1 grados) se utilizaron 3 binomios para localizar a 53 personas de las cuales, 12 de ellas aún estaban con vida. Además de la localización de personas, estos binomios también participan en la localización de drogas y explosivos donde juegan un papel fundamental para la preservación de la vida del personal de policía y seguridad de civiles[2].

Estos perros se enfrentan a tareas en condiciones extremas, por ejemplo, calor extremo, terreno irregular, fugas de gas, derrumbes súbitos, réplicas de sismos, entre muchos otros peligros. Tales condiciones exigen que los perros posean condiciones de personalidad y comportamiento muy específicas tal como se describe en el manual de International Search and Rescue(IRO) [3] las cuales incluyen obedecer al entrenador, tener mínima reacción ante ruidos fuertes (disparos), ser capaz de entrar a espacios reducidos, poseer gran agilidad y obedecer comandos de voz.

Debido a estas características la selección y entrenamiento de perros aptos para tareas de búsqueda y rescate es complejo y tardado (Figura 1). Si bien cualquier raza de perro puede realizar las pruebas de selección, solo ciertas razas de perros cumplen con las aptitudes físicas y emocionales necesarias y de estos, muy pocos ejemplares resultan debidamente capacitados. Otros factores que influyen en la aptitud del perro para estas tareas son la pureza de la raza o pedigrí y la crianza.

**Figura 1.** Perro realizando entrenamientos de simulación de zonas de desastre controladas.

Se estima que en México hay aproximadamente 300 perros (PByR) y forman parte de la sección canina de la Marina [4]. También, la UNAM cuenta con 29 binomios que pertenecen a su unidad K9 del programa de manejadores de perros de búsqueda y rescate, sin embargo, solo una pequeña parte de ellos están especializados para realizar tareas de localización de personas en áreas de desastre [5].

## Antecedentes

Recientemente ha habido un creciente interés en investigaciones que tienen como finalidad estimar el estado interno de diferentes animales, entre ellos los perros. Esto es, usar sensores embebidos en dispositivos de tipo vestible para registrar la reacción de un perro ante un estímulo externo durante la realización de una tarea para clasificar el desempeño del animal y estimar sus emociones ya sea negativas, positivas o neutras. La Figura 2 muestra un diagrama con la arquitectura típica de sistemas de adquisición de datos de manera remota y la posterior aplicación del algoritmo de aprendizaje de máquina.

Además de monitorizar en tiempo real, se trata de determinar si el perro es apto o no para entrenarse en tareas de búsqueda y rescate, asistencia o guardia. Estos

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

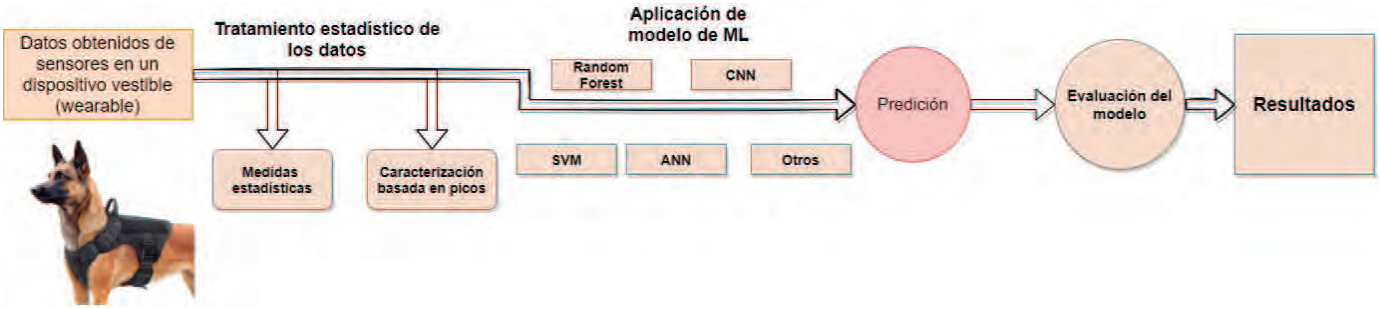



**Figura 2.** Modelo de Aprendizaje de Máquina básico para sistemas de tipo vestible (wearable) para perros.

sensores se colocan en chalecos, pecheras o bandas en las patas. En 2014, investigadores de la universidad de Carolina del Norte [6] (USA) utilizaron un chaleco colocado en el lomo del perro con los siguientes sensores: IMU (unidad de masa inercial), sensores de gas, ECG (muestreo de Electrocardiógrafo) y PPG (Fotopletismografía) para medir la oxigenación. Una imagen del chaleco se puede observar en la Figura 3. El dispositivo se comunica de forma inalámbrica para permitir al perro realizar actividades de manera cómoda.

Con la información de las IMU y utilizando algoritmos de aprendizaje de máquina en cascada, fue posible estudiar la cinemática del cuerpo del perro y clasificar su postura estática en cinco actividades (sentarse, pararse, echarse, parase en dos patas, comer de un plato en el suelo) y tres posturas dinámicas (caminar, subir escaleras, bajar rampas).

El método de aprendizaje automático siguió dos etapas: la primera etapa utilizando HMM (Modelos ocultos de Markov, autómatas que permiten modelar procesos estocásticos), luego un árbol de decisiones C4.5 y finalmente, una validación cruzada de 10 pliegues. En las mediciones participaron ocho perros de diferentes razas medianas y grandes.

Otra investigación encabezada por Brugarolas [7] de la Universidad de Carolina del Norte utilizó mediciones de diferentes electrodos comerciales en contraste con un prototipo para tener mediciones más exactas de las señales biométricas de perros.

Estos se han empleado en perros para monitorizar su ritmo cardíaco (HR) así como sus intervalos R-R (intervalo de tiempo entre dos pulsos del ritmo sinusoidal cardiaco).

Se ha encontrado que arreglos circulares colocados como pasadores cuyas puntas están cubiertas de un tipo de polímero conductivo reducen la impedancia entre el contacto del electrodo y la piel del perro, proporcionando mediciones más exactas y una caracterización de ECG y PPG in vivo. Utilizando esta información fue posible detectar una correlación precisa entre el ritmo cardíaco, la información de un acelerómetro y la respiración en intervalos R-R y una posición estática del perro.

En la investigación de Aich et al [9] se utilizaron diferentes técnicas de aprendizaje de máquina como Random Forest, SVC, KNN, Redes Bayesianas y ANN para la detección de actividad y emoción de perros de diferentes razas colocando IMUs en el cuello y la cola del perro. Esto debido a que existe una estrecha relación entra la emoción que siente un perro (positiva, negativa o neutra) dependiendo la dirección en la que ondea la cola al moverse.

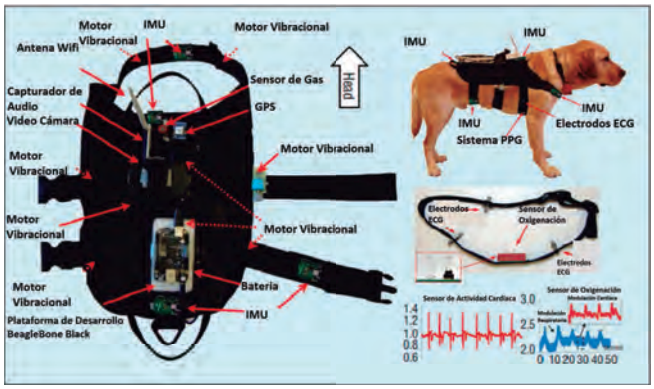



**Figura 3.** Wearable colocado en el lomo del perro, con sensores embebidos, ECG,PPG,IMU. Imagen Obtenida de [6].

También se utilizaron diferentes modalidades (posición y vídeo) para realizar una sincronización, pretratamiento y fusión de datos. Con esto fue posible relacionar los movimientos de la cola y pecho del perro con la actividad que este realiza y la emoción que manifiesta al realizarla. Se predijeron con éxito las siguientes actividades: caminar, sentarse, mantenerse estático, caminar a lado del instructor, comer, saltar y realizar búsqueda de objetos mediante el olfato.

La investigación de Hayati et al [9] realizó el análisis en tiempo y frecuencia de la marcha de perros de razas ágiles (galgos) colectando video y señales de IMU en 31 de ellos y después aplicando la Trasformada Rápida de Fourier(FFT) y la ondícula de Morlet (Morlet Wavelet). Se observó que los picos de aceleración corresponden a los golpes de las patas traseras y equivalen a quince veces el peso del galgo. En la Figura 4 se puede observar par-

te del análisis. De la misma manera fue posible realizar una descripción detallada en la cinemática del movimiento del perro desde que inicia en una posición de espera y se encontró un patrón característico en la aceleración dorsal-ventral de cada perro al realizar las pruebas.

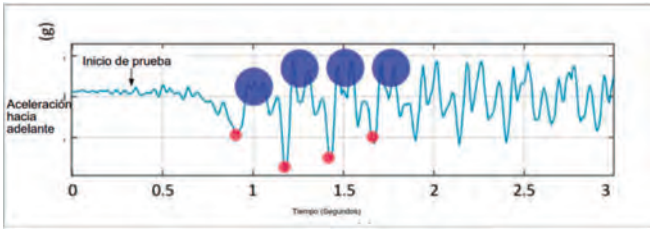



**Figura 4.** Aceleraciones experimentadas por perros galgos en ocho zancadas continuas. Los picos negativos corresponden a aceleración en las patas delanteras (en rojo) y las aceleraciones positivas en las patas traseras (en azul). Imagen tomada de [9].

En un segundo experimento fue posible observar como las condiciones del terreno afectan a la velocidad del perro, según la dirección en la que se encuentran en un circuito. Esto refuta lo documentado en la literatura acerca del peligro que enfrentan los perros al realizar giros en un circuito de competición en comparación a tener una trayectoria recta. También se encontró que no existe diferencia de velocidad ocasionada por el terreno al comparar pasto con arena, ambos materiales empleados con frecuencia en los circuitos. En [10] mediante el uso de un vestible con un GPS y un sensor de mordida capacitivo se demostró que es posible entrenar a un perro para enviar una señal de haber encontrado algo mordiendo un sensor. El entrenador recibe la alerta y ubicación mediante una interfaz en un teléfono móvil.

## Limitaciones

A pesar del creciente interés en el área, existen ciertas limitaciones que enfrentan los dispositivos desarrollados hasta el momento, principalmente relacionadas a la fiabilidad de las mediciones que en zonas de desastre pueden resultar afectadas por las condiciones extremas. Aunado a esto existen múltiples complicaciones que puede enfrentar un perro al utilizar dispositivos de gran tamaño pues puede reducir la movilidad del perro al entrar en espacios reducidos. Siendo esta última una de las preocupaciones principales de los entrenadores ya que los perros pueden quedarse atascados por el chaleco si este es muy voluminoso o que este pueda romperse y ocasionar una reacción inesperada en el perro y poner en peligro su bienestar.

Frecuentemente los perros realizan recorridos en zonas afectadas por un desastre en los cuales el perro se aleja del campo visual del entrenador y se pierde la interacción. En estas zonas las redes telefónicas móviles no están disponibles para las soluciones que implementan GPS para determinar la ubicación del perro. Para lidiar con este problema se puede implementar una red de nodos de sensores inalámbricos (WSN) colocados en múltiples perros de rescate de forma que se comuniquen (tipo malla) entre ellos y envíen la información a un nodo sumidero. Posteriormente se realiza el almacenamiento y procesamiento de la información en la nube [16]. También se ha propuesto implementar antenas móviles que funcionen como nodos sumideros para casos en los que no se tengan múltiples perros de rescate.

Las limitaciones principales encontradas en las investigaciones son las siguientes:

- Las condiciones extremas a las que se enfrentan los perros afectan a las comunicaciones inalámbricas (temperatura, grandes distancias, entrar en edificios derrumbados, saturación de redes).
- Los perros pueden presentar comportamientos extraños ante una situación real que puede no estar registrada previamente y presentar un estado interno del perro poco preciso.
- Los perros al realizar el recorrido de las zonas afectadas se alejan del campo visual del entrenador, debe existir una forma de comunicación entre el perro y el entrenador para enviar un mensaje de localización o de encontrarse en una situación de peligro.
- Debido a las limitaciones de capacidad de procesamiento en los chalecos, batería y tamaño, no es posible realizar el análisis de la información recolectada por los sensores de manera local por lo que debe ser enviada para su procesamiento. Esto genera complicaciones en el tiempo de funcionamiento del dispositivo.
- Las investigaciones en general utilizan una modalidad para describir el comportamiento de los perros. La mayoría se centra en la postura o biométricos.
- Para que la predicción del estado interno y la actividad del perro sea útil para el entrenador, es necesario que la información recopilada y mostrada en una interfaz móvil o de otro tipo sea lo más cercano al tiempo real. De está manera el entrenador podrá tomar decisiones precisas y monitorizar al perro para maximizar la eficiencia del binomio y garantizar su bienestar.

## Nuestro Trabajo

Nuestro equipo de trabajo formado por personal de Universidad Autónoma de Yucatán (UADY), la Universidad Autónoma de Tlaxcala (UATx), el Instituto Nacional de

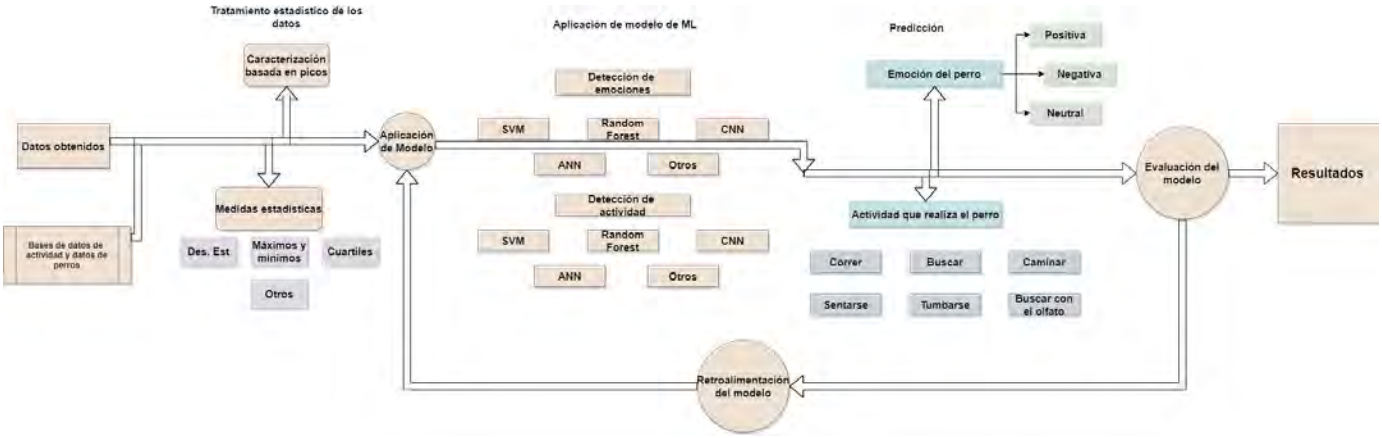



**Figura 5.** Modelo propuesto del proyecto.

Astrofísica, Óptica y Electrónica (INAOE) en colaboración con Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE) trabajamos en una propuesta para realizar el perfilado de los perros que aspiran a ser candidatos a perros de búsqueda y rescate. Utilizando técnicas de aprendizaje de máquina de manera multimodal sobre datos provenientes de sensores en un dispositivo vestible portado por un perro se pretende predecir el estado interno del perro cuando se enfrenta a estímulos externos típicos en una escena de búsqueda y rescate. También se quiere reconocer actividades físicas como caminar, sentarse o correr a partir de los datos de la pechera.

## Método Propuesto

La Figura 5 muestra un diagrama con los componentes y el funcionamiento del proyecto. Los datos provienen ya sea de la pechera, que corresponde a otra parte del proyecto, o de bases de datos disponibles. Las señales o modalidades seleccionadas para realizar el proyecto son postura o posición y biométricos (ECG,PPG) ya que se ha encontrado una relación clara entre estas dos modalidades y el estado interno del perro, sus emociones y la actividad que realiza. De manera similar a como los entrenadores usan distintos sentidos para interpretar el estado interno del perro, se espera que al incorporar varios canales de información nuestro sistema interprete automáticamente la emoción y actividad del perro.

Nuestras hipótesis son las siguientes:

- Se obtendrán los datos correspondientes a los sensores de posición (acelerómetros, giroscopios) y se podrá caracterizar y predecir la actividad que realiza el perro
- Se podrá monitorizar el estado interno del perro, su emoción ante un estímulo externo y cómo reacciona ante este
- Se determinará qué técnica de aprendizaje de máquina resulta efectiva para estimar la emoción y actividad del perro,
- Utilizando los criterios correspondientes a [5] se podrá sugerir si un perro es apto o no para ser entrenado en tareas de búsqueda y rescate.
- Se podrá proporcionar información sobre el estado del perro a los entrenadores que les permita tomar mejores decisiones al entrenar o trabajar en campo.

## Conclusiones

Existe una tendencia a utilizar dispositivos electrónicos vestibles para monitorear la actividad de perros, además, investigaciones recientes muestran que es posible usar Inteligencia Artificial para predecir emociones o reconocer actividades utilizando técnicas de aprendizaje de máquina sobre datos de movimiento, vocalización o biométricos obtenidos de tales dispositivos vestibles. En este artículo se discutieron algunos enfoques y se presentó un proyecto que aprovecha múltiples modalidades para realizar las predicciones debido a que se cuenta con mayor información para que los algoritmos computacionales realicen su tarea. De manera análoga a lo que sucede con los humanos, se espera que la fusión de información auditiva con información de movimiento e información biométrica obtenga mejores resultados que usar una sola fuente. El enfoque estuvo puesto en PByR (perros de búsqueda y rescate) pero los resultados bien podrían aplicarse a perros de asistencia y compañía. De manera comercial existen soluciones que permiten la monitorización de animales pero aún queda mucho por explorar en está área debido a las múltiples actividades e interacciones que tenemos los seres humanos con los perros. ✳

## REFERENCIAS

1. Gobierno de México. (2017). Sismología de México. Recuperado el 19 de Mayo de 2021 de https://www.sgm.gob.mx/Web/MuseoVirtual/Riesgos-geologicos/Sismologia-de-Mexico.html.

2. JCB Televisa News. (2017). Conoce a los perros rescatistas de la Marina que salvaron vidas en CDMX y Oaxaca. Recuperado 20 Mayo 2021 de https://www.ngenespanol.com/naturaleza/informacion-sobre-los-perros-de-rescate/amp/.

3. International Search and Rescue Dog Organization. (2019). International Trial Rules for Search and Rescue Dogs Tests. Manual de competición. Salburgo, Austria. International Rescue Organization. Recuperado el 21 de Mayo de 2021 de http://www.fci.be/medias/SAU-REG-IPO-en-2019-9050.pdf.

4. Delgado, E. (2017). ¿Qué ha pasado con los perros rescatistas del 19-S? *Milenio*. Recuperado 25 de mayo 2021 de https://www.milenio.com/politica/comunidad/perros-rescatistas-del-19-s-que-ha-pasado-con-ellos.

5. Carrillo, M. (2021). Unidad Canina K9: así es como la UNAM entrena binomios perro-humano para rescate en desastres. Recuperado el 24 de Mayo del 2021 de https://www.milenio.com/politica/comunidad/perros-rescatistas-del-19-s-que-ha-pasado-con-ellos.

6. Bozkurt, A., Roberts, D., Sherman, B., Brugarolas, R., Mealin, S., Majikes, J., Yang, P., y Loftin, R. (2014). Toward CyberEnhanced Working Dogs for search and rescue. *Cyborg Intelligence*, 29(6), 32-49.

7. Brugarolas, R., Latif, T., Dieffenderfer, J., Walker, K., Yuschak, S., Sherman, B., Roberts, D., Bozkurt, A. (2016). Wearable Heart Rate Sensor Systems for Wireless Canine Health Monitoring. *IEEE Sensors Journal*, 3454-3464.

8. Aich, S., Chakraborty, S., Sim, J., Jang, D., y Kim, H. (2019). The Design of an Automated System for the Analysis of the Activity and Emotional Patterns of Dogs with Wearable Sensors Using Machine Learning. *Applied Sciences 2019*, 9, 4938.

9. Hayati, H., Mahdavi, F., y Eager, D. (2019). Analysis of Agile Canine Gait Characteristics Using Accelerometry. *Sensors*, 19(20), 4379.

10. Ferworn, A., Waismark, B., y Scanlan, M. (2015). CAT 360 — Canine augmented technology 360-degree video system. *IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, 1-4.


## SOBRE LOS AUTORES

**Rodrigo Chi Ozalde** es Ingeniero Mecatrónico por la universidad Anáhuac Mayab. Actualmente cursa como estudiante de tiempo completo de la maestría en ciencias de la computación. Entre sus interés están la aplicación de algoritmos de aprendizaje de máquinas para resolución de problemas relacionados con seres vivos y su comportamiento. Entre sus pasatiempos favoritos se encuentran la música y los idiomas.

**Humberto Pérez-Espinosa** recibió el grado de Licenciado en Ciencias Computacionales en la Benemérita Universidad Autónoma de Puebla (BUAP) en 2004, el grado de Maestro en Ciencias Computacionales en 2006, y el grado de Doctor en Ciencias Computacionales en el Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE) en 2013. Sus intereses de investigación incluyen el aprendizaje automático, computación afectiva, computación paralingüística y el análisis inteligente de audio. Actualmente es catedrático CONACyT en la Unidad de Transferencia Tecnológica del CICESE y miembro del Sistema Nacional de Investigadores nivel uno.

**Jorge Alberto Ríos Martínez** Obtuvo los grados de Licenciado en Ciencias de la Computación y Maestro en Ciencias Matemáticas en la UADY, en 2002 y 2007, respectivamente. Obtuvo el grado de Doctor en Informática y Matemáticas aplicadas en la Universidad de Grenoble, Francia en 2013. Desde el 2013, se desempeña como profesor de tiempo completo en la Facultad de Matemáticas de la UADY. Sus intereses de investigación se centran en la robótica móvil y en aplicaciones del aprendizaje de máquina.

Imagen tomada de: https://www.ibm.com/mx-es/analytics/machine-learning.

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

ARTÍCULO ACEPTADO

# Inteligencia Artificial Explicable XAI

Tonantzin Marcayda Guerrero Velázquez y Juan Humberto Sossa Azuela

## Introducción

Actualmente, el uso de la Inteligencia Artificial (IA) y del Aprendizaje Automático o Machine Learning (ML), están innegablemente presentes en la mayoría de los aspectos de nuestras vidas.

La IA trata de emular el comportamiento humano y de hacer que las máquinas razonen y actúen lo más parecido posible a los seres humanos. Por otro lado, el Machine Learning es una rama de la IA cuyo objetivo es aprender a reconocer patrones dentro de los datos que le son suministrados, con la finalidad de realizar predicciones con base en esa información. Este proceso se realiza de manera automática, es decir, sin necesidad de que un ser humano interfiera en el proceso de aprendizaje para llegar a los resultados obtenidos por este tipo de modelos.

El desarrollo de una gran cantidad de trabajos de investigación dentro de estas áreas, así como el que existan computadoras cada vez más potentes y acceso a una gran cantidad de información, ha permitido que cada día se generen nuevos métodos y técnicas que logran resultados increíbles, inclusive más allá que los obtenidos por un ser humano [1]. Debido a esto, hoy en día, muchos problemas del mundo real en diferentes áreas tales como: las finanzas, la venta de productos y servicios, la medicina, la atención médica, el manejo de vehículos autónomos, entre otros, se empezaron a abordar aplicando modelos de aprendizaje automático [1]. Los resultados obtenidos por estos modelos, dado que influyen de manera directa en la vida de las personas, se tornan realmente críticos.

Los modelos de aprendizaje automático también permiten la creación de herramientas que nos facilitan nuestro día a día y que nos ayudan en tareas tan simples como realizar una búsqueda en la web. A continuación, se listan algunos ejemplos de ello:

- *Reconocimiento de voz.* En nuestro día a día solemos utilizarlo para darle órdenes a nuestros dispositivos inteligentes, ya sea a nuestro Smartphone, al famoso Chromecast de Google, o Alexa de Amazon.

- *Motores de Búsqueda.* Frecuentemente utilizamos buscadores cuando navegamos en la web, los cuáles nos permiten encontrar una inmensa cantidad de información referente a algo que es de nuestro interés. El más utilizado a nivel nacional según las estadísticas de StatCounter [2] es Google, seguido de Bing y Yahoo.
- *Recomendaciones automáticas.* Los servicios de *streaming* que comúnmente utilizamos, tales como Netflix o Spotify, nos ofrecen este tipo de recomendaciones con base en nuestros gustos y preferencias, con los cuáles estos sistemas aprenden a conocernos más y mejor para así ofrecernos un mejor servicio al cliente.
- *Ventas de productos y Marketing.* Diversas plataformas de ventas de productos por Internet hacen uso de la IA para crear campañas de marketing con las cuáles dirigen sus esfuerzos en influenciar a las personas a comprar un producto u otro. El análisis del comportamiento del usuario en estas plataformas permite que los algoritmos aprendan a identificar los gustos de los clientes y sus preferencias, y con esto hacer recomendaciones personalizadas y dentro de un segmento en específico.
- *Traducción automática.* Dentro de nuestras múltiples tareas, seguramente en alguna ocasión hemos tenido la necesidad de utilizar este tipo de herramientas. Los traductores automáticos ayudan a las personas a tener acceso a una mayor cantidad de información, ya que, aunque esa información no esté en su idioma las herramientas les ayudarán a entenderla. Dentro de esta área hay una gran cantidad de investigación que se enfoca en que estos traductores cada vez den resultados más precisos y que aprendan no solo a simplemente traducir, si no a interpretar sus traducciones. El reto es conseguir que las traducciones sean naturales y coherentes. Un ejemplo de este tipo de herramientas ampliamente utilizado en Google Traductor.
- *Conducción autónoma de vehículos automotores.* Grandes empresas tales como Tesla, Uber o Google, entre otras, han enfocado sus investigaciones hacia este rubro, con la finalidad de algún día tener vehículos totalmente autónomos.

**Dada su alta precisión, el Machine Learning es utilizado frecuentemente en la industria y en aplicaciones de la vida cotidiana.**

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

El que estos modelos entreguen tan buenos resultados hace que con mayor frecuencia sean utilizados y adaptados tanto en la industria como en nuestra vida diaria. Sin embargo, el que tengan tan alta efectividad y precisión, nos puede llevar a reflexionar y a preguntarnos ¿Cómo es que lo hacen? ¿Cómo es que llegan a ser tan precisos? Actualmente los modelos de aprendizaje automático manejan enormes cantidades de información para obtener sus resultados y llegar a una predicción óptima. La mayoría de estos modelos utilizan algoritmos de aprendizaje profundo, los cuáles son muy difíciles de interpretar debido a la gran cantidad de operaciones que realizan para llegar a sus resultados.

Este tipo de modelos suelen llamarse Modelos de Caja Negra, nombrados así por que únicamente se conocen los resultados obtenidos, pero no se explica la manera de cómo es que se llegaron a ellos. Es decir, el usuario no conoce sus componentes internos o la lógica que lo llevo a obtener sus resultados. Estos modelos de caja negra resultan ser complejos modelos de aprendizaje automático, de los cuáles es muy difícil obtener una explicación que permita entender el porqué de sus predicciones. Está falta de explicabilidad [3], lleva a que los usuarios generen cierta reticencia a confiar en ellos, especialmente cuando se utilizan en ámbitos cruciales o cuando manejan información sensible, por ejemplo, para la autorización de algún crédito bancario, el estado de salud de un paciente, el manejo de vehículos autónomos, entre otros.

## Inteligencia Artificial Explicable

Hoy en día, la Inteligencia Artificial Explicable (XAI por sus siglas en inglés), se ha vuelto un área de mucho interés dentro del campo del aprendizaje automático. Lo que se busca con la XAI es mejorar el entendimiento humano sobre los modelos artificiales, así como tratar de justificar sus decisiones.

La DARPA (Defense Advanced Research Projects Agency, por sus siglas en inglés) define a la XAI como aquellos sistemas de Inteligencia Artificial que pueden explicar su proceso de toma de decisiones a un ser humano, caracterizar sus fortalezas y sus debilidades, así como tener la capacidad de transmitir un entendimiento de cómo se comportarán en el futuro [4].

En el contexto de la Inteligencia Artificial, la explicabilidad se refiere a cualquier acción o proceso que se lleve a cabo con la intención de aclarar el proceso de decisión de un modelo. La mayoría de las veces el concepto de explicabilidad se usa de manera indistinta con el concepto de interpretabilidad. Sin embargo, son conceptos que involucran enfoques distintos. La interpretabilidad se refiere al nivel en el que un modelo dado tiene sentido para un ser humano, y este concepto también puede expresarse como la transparencia de un modelo. Un modelo es considerado transparente si por sí mismo es entendible, por ejemplo, un árbol de decisiones o un clasificador basado en reglas.

En [4], DARPA describe que existe una aparente relación presente entre el desempeño de un modelo de aprendizaje automático y su explicabilidad ya que aquellos métodos que tienen un mayor rendimiento, tales como los modelos de aprendizaje profundo, resultan ser los menos explicables. Por otro lado, aquellos modelos que resultan más explicables, por ejemplo, un árbol de decisión, son los menos precisos. En la Figura 1 (que es una versión adaptada de la Fig. 1 perteneciente al artículo citado en [4]), se muestra la relación descrita anteriormente.

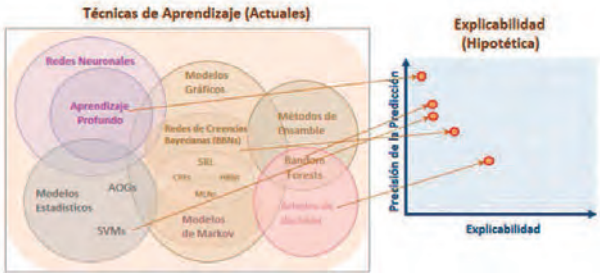



**Figura 1.** Rendimiento del aprendizaje vs. Nivel de explicabilidad.

**La XAI busca mejorar el entendimiento humano sobre los modelos artificiales así como el tratar de justificar sus decisiones.**

Con el uso de los algoritmos de IA en áreas tan críticas como lo son la salud, las finanzas, el manejo de vehículos autónomos, entre otros, la necesidad de explicar el porqué de sus decisiones resulta importante por razones éticas, de seguridad y legales [5]. Así mismo, la explicabilidad ayuda a entender el trasfondo de las explicaciones, de tal manera que se puedan mostrar aquellas características que resulten más importantes para el modelo al momento de llevar a cabo una predicción o tomar una decisión. La explicabilidad también puede ayudar en la mejora del modelo, ya que, al mostrar las características de mayor relevancia, se podría alimentar al modelo con esa información para así buscar mejorar sus predicciones. Debido a la necesidad de analizar las decisiones que toman los modelos de aprendizaje automático, han surgido una gran variedad de métodos de explicación. En [6] los autores proponen tres clases principales para clasificarlos. En la primera clase se encuentran aquellos

métodos basados en la extracción de reglas, cuyo objetivo es aproximar el proceso de toma de decisiones del modelo utilizando sus entradas y salidas. La segunda clase corresponde a los llamados métodos de atribución, los cuáles miden cuánto afecta el cambio de las entradas o de sus componentes internos al rendimiento del modelo. En la última clase se encuentran los métodos llamados intrínsecos, en donde el objetivo es mejorar la interpretabilidad de lo métodos implementando métodos derivados de la arquitectura del modelo.

Dentro de las diversas técnicas existentes que proveen una explicación para los modelos de aprendizaje, se encuentran los métodos LIME (Local Interpretable Model-Agnostic Explanations) y RISE (Randomized input sampling for explanation of black-box models).

A continuación, daremos una breve explicación de cómo es que estos métodos obtienen una explicación de las predicciones realizadas por un clasificador de imágenes. Por una lado LIME crea un conjunto de instancias que son el resultado de perturbar la imagen de entrada al dividirla en componentes interpretables (superpixeles), con la finalidad de obtener una probabilidad de pertenencia para cada una de estas instancias perturbadas. Una vez hecho lo anterior, LIME genera una explicación visual basada en la clasificación de esos nuevos datos perturbados, con lo cual, se obtiene un área de la imagen de entrada la cuál denota en lo que un modelo se fijó para llevar a cabo su predicción [7]. Por otro lado, lo que hace RISE [8] es producir un mapa de calor o de saliencia, en el cual se muestran aquellas partes de la imagen de entrada que son más importantes para la predicción hecha por el clasificador. Este mapa de calor se obtiene mediante la generación de máscaras aleatorias y superponiéndolas en la imagen original. Después, estas versiones de la imagen original con las máscaras superpuestas, alimentan a la red neuronal para observar los cambios que ocurren en la salida de la red. Cuando este proceso se repite muchas veces, es posible identificar qué características de la imagen de entrada son más importantes para la predicción realizada por el modelo.

Sin embargo, existe una gran desventaja en los métodos actuales que ofrecen explicaciones visuales la cual se refiere a la subjetividad de los resultados, dado que éstos se encuentran sujetos a la interpretación de quienes los observan. Como consecuencia, se puede llegar a cuestionar la fiabilidad del método de explicación.

En el Centro de Investigación en Computación del IPN, en el laboratorio de Robótica y Mecatrónica, hemos desarrollado experimentaciones orientadas hacia el área de la Inteligencia Artificial Explicable, específicamente se diseñó un nuevo método de explicabilidad basado en la clasificación de regiones útiles en una imagen, con el cuál se logra superar el problema descrito anteriormente: la subjetividad de los resultados obtenidos. Para conocer más detalles sobre nuestro trabajo, se pueden consultar las referencias citadas en [9] y [10].

**Existe una gran desventaja en los métodos actuales que ofrecen explicaciones visuales: la subjetividad de los resultados.**

## Trabajo de investigación desarrollado

Con el desarrollo de este nuevo método de explicabilidad [9], se creó una solución para superar el problema de la subjetividad presente en los resultados obtenidos por los métodos de explicación actuales. La solución propuesta consiste en la creación de una explicación visual de la predicción, basada en la caracterización de ciertas regiones de la imagen de acuerdo con su importancia para la predicción del modelo. La principal contribución de este trabajo radica en la creación de un novedoso método de explicabilidad con el cual se logran obtener explicaciones no subjetivas; esto se logra de dos maneras: 1) No existen parámetros de configuración para el algoritmo, con lo cuál se asegura que no depende de quien lo implemente y 2) El resultado de la explicación es claro y fácil de entender así como intuitivo, debido a las categorías propuestas para clasificar a cada una de las regiones útiles resultantes. El principal objetivo del método propuesto es identificar aquellas regiones de la imagen de entrada que sean más importantes para la predicción del clasificador, para posteriormente categorizarlas como significativas, relevantes o fútiles. Cada una de estas regiones son resaltadas con los colores verde, amarillo y rojo respectivamente, con lo cual se logra obtener una explicación visual. Este resultado se logra primero realizando una búsqueda selectiva de regiones dentro de la imagen de entrada con lo cual se obtiene un conjunto de regiones candidatas; llamadas así porque estas regiones podrían resultar importantes para el clasificador, pero realmente aún no se sabe si lo son o no. Por lo tanto, este conjunto de regiones candidatas es evaluado utilizando el mismo clasificador y posteriormente sometidas a un análisis estadístico para así seleccionar las regiones más importantes, a las cuáles se les llama regiones útiles. Finalmente, estas regiones útiles se categorizan y colorean como significativas, relevantes o fútiles. Los detalles de este proceso se describen en nuestro trabajo citado en [9]. En la Figura 2 se muestran ejemplos de regiones candidatas encontradas por el algoritmo de búsqueda selectiva, para algunas imágenes del conjunto de datos

*Dogs. vs Cats* tomado de *Kaggle*, clasificadas como A) Cocker spaniel, B) Egyptian cat, C) Egyptian cat y D) Norwegian elkhound por el modelo InceptioResnet.

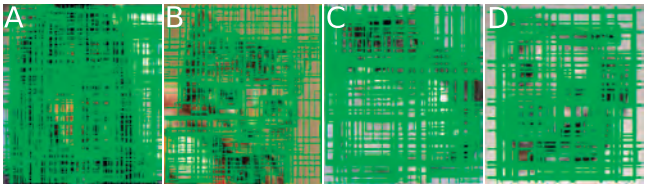



**Figura 2.** Regiones resultantes al aplicar el algoritmo de búsqueda selectiva [9].

En la Figura 3 se muestra el conjunto reducido de las regiones candidatas mostradas en la Figura 2. Este conjunto se obtiene al evaluar con el mismo clasificador cada una de las regiones candidatas, obteniendo así aquellas que pertenecen a la clase que se desea explicar, y posteriormente sometiéndolas a un análisis estadístico con lo cual se seleccionan las regiones realmente útiles.

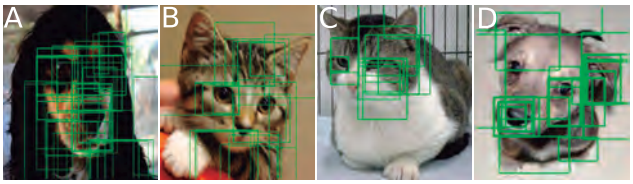



**Figura 3.** Conjunto de regiones útiles resultantes para los ejemplos mostrados en la Figura 1 [9].

Finalmente, cada una de estas regiones útiles será evaluada para clasificarla dentro de una de las tres categorías propuestas: significativa, relevante y fútil. Como está implícito en los nombres, cada una de estas categorías se refiere al nivel de importancia que cada una de ellas tiene para la toma de decisiones del clasificador. En la figura 4, finalmente se puede observar la explicación visual resultante para la clase predicha por el modelo en cada uno de los ejemplos mostrados en la Figura 2. Para consultar los detalles de este proceso de caracterización, favor de consultar la referencia citada en [9].

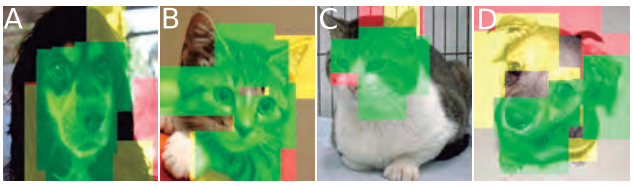



**Figura 4.** Explicación visual resultante para las imágenes de la Figura 2 [9].

A continuación, presentamos una breve descripción de algunos de los resultados obtenidos al aplicar el método de explicabilidad propuesto. Para ampliar esta información favor de consulta la referencia citada en [9].

## Resultados y Discusión

En la Figura 5, se pueden observar las explicaciones resultantes para tres imágenes que representan distintas escenas. En el primer ejemplo, el modelo clasificó a esta imagen como clase *muelle* con una probabilidad igual a 0.728. Se puede apreciar que las regiones obtenidas y marcadas por el método como una explicación de la predicción del modelo nos hacen mucho sentido, ya que con solo mirar estas regiones es posible decir que la clase predicha es la correcta. Para el segundo y tercer ejemplo la lógica es similar, y se observa que las explicaciones resultantes también engloban aquellas regiones que como seres humanos seguramente marcaríamos como importantes. La segunda imagen fue clasificada como *acueducto* con una probabilidad de pertenencia igual a 0.935, y la tercer imagen fue clasificada como *paso peatonal* con una probabilidad de pertenencia igual a 0.378.

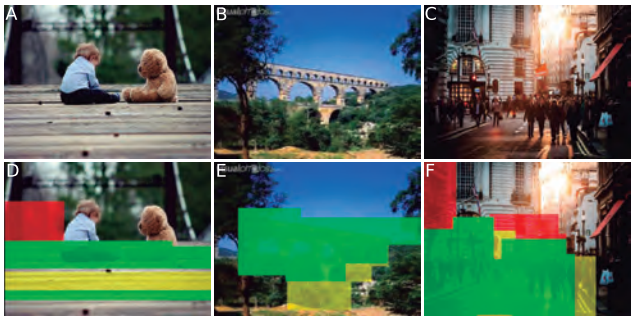



**Figura 5.** Ejemplo de diferentes imágenes que representan una escena y su explicación obtenida con el método propuesto [9].

La relevancia del método radica también en la importancia de que nos permite conocer la confiabilidad de las predicciones dadas por un modelo ya que, por definición, ningún modelo es perfecto, por lo tanto, cuando se llega a obtener una predicción incorrecta, sería de mucha ayuda conocer el motivo y con ello poder mejorar el modelo.

En la Figura 6, se pueden apreciar tres ejemplos de imágenes que a simple vista podríamos decir que pertenecen a la clase gato, sin embargo, el modelo las clasifica a clases totalmente distintas, es decir, identifica como clase principal otra clase que nosotros no consideraríamos como tal.

**Figura 6.** Imágenes clasificadas por el modelo InceptionResnet como A) Cesta de compra, B) Edredón y C) Shoji [9].

Entonces, gracias a la explicación visual generada por nuestro método es posible saber exactamente cómo es que el modelo toma su decisión de clasificación, e identificar las regiones de interés para la asignación de estas clases. Esto se puede observar en la Figura 7.

**Figura 7.** Explicación visual resultante para las imágenes de la Figura 6 [9].

Como se mencionó anteriormente, nuestro método resuelve uno de los problemas que están fuertemente presentes en otros métodos de explicación visual tales como LIME y RISE, que es la subjetividad de sus resultados.

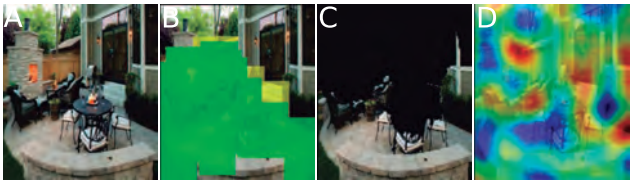



**Figura 8.** Explicaciones para una misma imagen usando distintos métodos. A) Imagen original, B) Método propuesto, C) LIME y D) RISE [9].

Como se aprecia en la Figura 8 (B), nuestro método supera este problema proporcionando resultados bien definidos de acuerdo con las diferentes categorías propuestas: significativa, relevante y fútil. En la Figura 8 (A) se muestra la imagen original la cuál fue predicha como clase patio por el modelo Resnet50Places365, con una probabilidad de pertenencia igual a 0.685. En las Figuras 8 (C) Y (D), se muestran las explicaciones resultantes de los métodos LIME y RISE respectivamente. En el caso de la explicación proporcionada por LIME, aunque es parecida a la obtenida por nuestro método, se observa que esta no es una región bien defina o delimitada (región no negra) ya que muestra un área no uniforme que incluye ciertas regiones de la imagen que podrían resultar difíciles de identificar y por lo tanto se pueden llegar a conclusiones diferentes para cada observador que analice los resultados. Por otro lado, RISE genera una explicación muy diferente y que podría generar una mayor confusión ya que las regiones de mayor interés son aquellas marcadas en color rojo y amarillo y como se puede ver, estas se encuentras dispersas por toda la imagen. Otro punto a considerar, es que se aprecia que existen regiones que se dejan fuera de la explicación y que podrían llegar a tener un peso importante para la predicción, por ejemplo, algunas de las sillas y mesas del mobiliario o parte del suelo que no ha sido marcado; o por el contrario, se llegan a considerar regiones que no tiene una clara relevancia en la predicción, por ejemplo la ventana, ya que aunque forme parte de la imagen una ventana puede aparecer en diferentes tipos de imágenes que pueden pertenecer a un tipo de clase diferente a la de patio. También se nota que a diferencia del método propuesto las explicaciones obtenidas con LIME y RISE no son tan claras y además son subjetivas, es decir, la interpretación puede ser diferente según el observador. Además, estos métodos también requieren una configuración previa de parámetros de los cuales depende el resultado obtenido.

Usando nuestro método también es posible obtener una explicación no solo para la clase principal predicha por el modelo, sino que asimismo se pueden explicar clases que tengan una probabilidad de pertenencia ligeramente menor. Por ejemplo, en la Figura 9 (A) se muestra la explicación para la clase principal predicha por el modelo clasificador como Weimaraner (raza de perro), y en las imágenes de la Figura 9 (B) y (C) las explicaciones para dos clases diferentes, cortina de baño y gato siamés, las cuáles también están presentes en la imagen.

**Figura 9.** Explicaciones visuales para múltiples clases presentes en una imagen [9].

Para verificar la validez de nuestro método, se llevó a cabo un proceso de comparación entre las regiones útiles obtenidas por el algoritmo propuesto y las regiones de interés seleccionadas por una persona dentro de una imagen, las cuales representan las regiones más importantes que determinan que una imagen pertenece a una clase u otra. Para hacer este estudio comparativo se utilizó el conjunto de datos COCO, el cual contiene imágenes etiquetadas de acuerdo con el o los objetos que se encuentren en ellas, y en las cuales una persona marca

un cuadro delimitador alrededor del objeto el cuál determina la región de interés. Esta región se usará como un indicador que nos ayudará a determinar si nuestro método funciona de manera ideal, ya que lo que la persona marco como región de interés para identificar un objeto, sería lo que idealmente una red neuronal artificial debería de considerar para clasificar una imagen.

En la Figura 10, se muestran cuatro ejemplos de imágenes pertenecientes al conjunto de datos COCO, las cuales han sido marcadas y clasificadas por un ser humano, y cuyos resultados serán comparados con las explicaciones obtenidas por nuestro método de explicación.

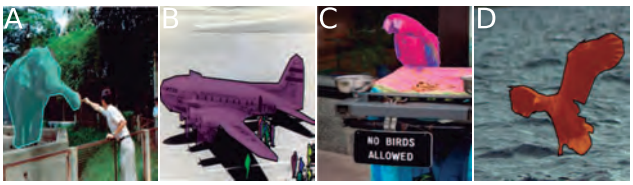



**Figura 10.** Ejemplo de imágenes pertenecientes al conjunto de datos COCO [9].

En la Figura 11 mostramos los resultados de las explicaciones resultantes aplicando nuestro método a las imágenes de la Figura 10. Es posible observar que la región útil encontrada por nuestro algoritmo rodea efectivamente toda la región marcada por el ser humano. Estos resultados coinciden bastante bien con los que se esperaría de un buen algoritmo de explicación visual.

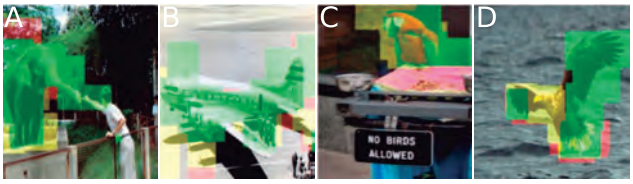



**Figura 11.** Resultado de la explicación visual de las imágenes de la Figura 10 [9].

## Conclusiones

La Inteligencia Artificial Explicable (XAI) se ha convertido en un área de gran interés, debido a que ha surgido la necesidad y el interés de obtener una explicación de las predicciones de los modelos de aprendizaje automático, ya que estos toman decisiones por nosotros y con mayor frecuencia se encuentran presentes en nuestra vida cotidiana. Con las técnicas de explicabilidad, se busca crear confianza en estos modelos ya que al impactar directamente en la vida de las personas en necesario justificar sus decisiones y tener un claro entendimiento de lo que están realizando. De esta manera se podrían verificar él porque fallan o si es que existe algún sesgo en la información proporcionada que propicie un mal desempeño del modelo. La principal aportación de nuestro método propuesto radica en que se ofrecen explicaciones no subjetivas, lo cual permite que la interpretación del resultado y la explicación siempre sea la misma, además de que resulta fácil de entender por cualquier persona. Actualmente nuestro método se enfoca en la generación de explicaciones visuales, sin embargo, se está trabajando para que su utilidad no este limitada solamente a clasificadores de imágenes, y pueda ser utilizado con otro tipo de datos y modelos. Sin duda alguna el desarrollo de investigación orientada al desarrollo de nuevos métodos y técnicas que mejoren el entendimiento de los modelos de aprendizaje profundo seguirá siendo un área de mucho interés e impacto, y en la cual hay un futuro muy prometedor. ✷

**Agradecimientos.** Tonantzin Guerrrero agradece a CONACyT la beca otorgada para realizar sus estududios de doctorado. Los autores agradecen al Instituto Politécnico Nacional por el apoyo económico otorgado en los proyectos SIP 20200630 y 20210788, y al CONACYT en los proyectos 65 (Fronteras de la Ciencia) y 6005 (FORDECYT-PRONACES).

## REFERENCIAS

1. Linardatos, P., Papastefanopoulos, V., y Kotsiantis, S. (2021). Explainable AI: A Review of Machine Learning Interpretability Methods". *Entropy*, 23(1), 18.

2. Statcounter GlobalStats (2020). Search Engine Market Share Mexico. Recuperado de Abril 2020 a Abril 2021, de `https://gs.statcounter.com/search-engine-market-share/all/mexico`.

3. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., y Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5), 1-42.

4. Gunning, D., y Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI Magazine*, 40(2), 44-58.

5. Das, A., y Rad, P. (2020). Opportunities and challenges in explainable artificial intelligence (xai): A survey. arXiv preprint arXiv:2006.11371.

6. Ras, G., van Gerven, M., y Haselager, P. (2018). Explanation methods in deep learning: Users, values, concerns, and challenges. En Escalante H. et al. (Eds.) *Explainable and Interpretable Models in Computer Vision and Machine Learning*. Springer, Cham. 19-36.

7. Ribeiro, M. T., Singh, S., y Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. En Proc. *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco California USA. 1135-1144.

8. Petsiuk, V., Das, A., y Saenko, K. (2018). Rise: Randomized input sampling for explanation of black-box models. arXiv preprint arXiv:1806.07421.

9. Velázquez, T. M. G., y Azuela, J. H. S. (2021). New explainability method based on the classification of useful regions in an image. *Computación y Sistemas*, 25(4).

10. Velázquez, T. M. G., y Azuela, J. H. S. (2020). Explicación visual de la predicción de un clasificador de imágenes. *Journal Research in Computing Science*, 149(11), 211-221.

## SOBRE LOS AUTORES

**Tonantzin Marcayda Guerrero Velázquez** es estudiante de doctorado en el Centro de Investigación en Computación del IPN. Sus intereses de investigación incluyen tópicos Inteligencia Artificial, técnicas y métodos de Inteligencia Artificial Explicable (XAI), Machine Learning y Data Science.

**Juan Humberto Sossa Azuela** es Ingeniero en Comunicaciones y Electrónica de la Universidad de Guadalajara, Maestro en Ciencias en Ingeniería Eléctrica por el CINVESTAV-IPN y Doctor en Informática por el Instituto Nacional Politécnico de Grenoble, Francia. Es profesor Titular "C" en el Instituto Politécnico Nacional en donde realiza investigación en Inteligencia Artificial, Aprendizaje para Máquinas y Robótica.

## *Inteligencia Artificial Explicable.*

La Inteligencia Artificial Explicable tiene como objetivo crear un conjunto de técnicas de aprendizaje automático que:

- Produzcan modelos más explicables, manteniendo un alto nivel de rendimiento de aprendizaje (precisión de predicción); y
- Permitan que los usuarios humanos comprendan, confíen adecuadamente y administren de manera efectiva la generación emergente de socios con Inteligencia Artificial.

Imagen tomada de: `https://www.iic.uam.es/innovacion/inteligencia-artificial-explicable-y-confiable/`.

ARTÍCULO ACEPTADO

# Una exploración sobre el aprendizaje automático simplificado: Generalización a partir de algunos ejemplos

Alejandro Hernández-Herrera, Elsa Rubio Espino, Rogelio Álvarez Vargas y Victor Hugo Ponce Ponce

## La importancia del aprendizaje de máquina utilizando pocos ejemplos de entrenamiento

Los seres humanos pueden aprender cosas nuevas con un pequeño conjunto de ejemplos. Cuando se les presentan estímulos, los humanos parecen ser capaces de comprender nuevos conceptos rápidamente y luego reconocer variaciones de esos conceptos en el futuro. Un niño puede aprender a reconocer a un perro a partir de una sola imagen, pero actualmente, un sistema de aprendizaje automático necesita muchos ejemplos para aprender las características de un perro y reconocerlas en el futuro.

El aprendizaje automático ha tenido mucho éxito en una variedad de tareas, como clasificación y búsqueda web, así como reconocimiento de imágenes y voz. Sin embargo, a menudo estos modelos requieren una gran cantidad de datos (ejemplos) utilizados para aprender. La motivación principal detrás de ésta modalidad de aprendizaje automático simplificado, es entrenar un modelo con muy pocos ejemplos que sea capaz de generalizar a categorías desconocidas sin un re-entrenamiento extenso.

En el aprendizaje asociativo estándar, un animal debe experimentar repetidamente una serie de conjugaciones entre un estímulo y una consecuencia antes de que se aprenda por completo un emparejamiento de estímulos en particular. Por lo tanto, el aprendizaje es inevitablemente incremental. Sin embargo, los animales a veces encuentran resultados que nunca antes habían experimentado y de los cuales es necesario aprender rápidamente para sobrevivir. En tales casos, los animales pueden aprender sobre la base de una sola exposición a un emparejamiento de estímulos, una situación descrita en la literatura como *aprendizaje simplificado* o *One-shot learning* [1].

Aprender rápidamente es un sello distintivo de la inteligencia humana, ya sea que implique reconocer objetos a partir de unos pocos ejemplos o aprender rápidamente nuevas habilidades después de unos minutos de experiencia. Los sistemas inteligentes artificiales deberían ser capaces de hacer lo mismo, aprender y adaptarse rápidamente a partir de unos pocos ejemplos, y seguir adaptándose a medida que haya más datos disponibles. Este tipo de aprendizaje rápido y flexible es un desafío, ya que el sistema debe integrar su experiencia previa con una pequeña cantidad de información nueva, evitando el sobreajuste a los nuevos datos. Además, la forma de experiencia previa y los nuevos datos dependerán de la tarea (ver Figura 1).

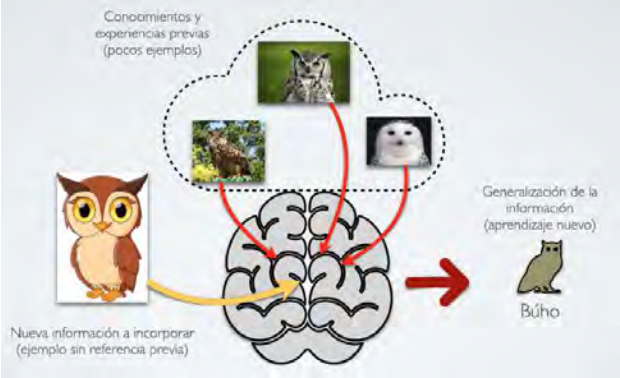



**Figura 1.** Aprendizaje de nueva información a partir de pocos conocimientos previos.

## Aprendizaje automático simplificado

Los modelos de aprendizaje automático profundo, basados en redes neuronales artificiales, han logrado un gran éxito en las tareas de reconocimiento visual [2]. Sin embargo, estos modelos de *aprendizaje supervisado*, es decir, en donde los resultados que se desean obtener del modelo son conocidos previamente y son utilizados para guiar su aprendizaje, necesitan grandes cantidades de *datos etiquetados* (datos conocidos previamente) y muchas iteraciones para entrenar su gran número de parámetros. Esto limita severamente su escalabilidad a nuevas clases de objetos emergentes (por ejemplo, nuevos dispositivos de consumo) o singulares (por ejemplo, animales extintos o peculiares) donde es posible que muchas imágenes etiquetadas simplemente nunca existan. Por el contrario, los seres humanos son muy buenos para reconocer objetos con muy poca supervisión directa, o ninguna, es decir, con un aprendizaje simplificado de pocas muestras [3] o uno de cero muestras [4].

Por ejemplo, los niños no tienen problemas para generalizar el concepto de "cebra" a partir de una sola imagen en un libro, o escuchar su descripción como si fuera un caballo rayado.

Motivado por el fracaso de los métodos convencionales de aprendizaje profundo para funcionar correctamen-

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

te en uno o algunos ejemplos por clase, e inspirado por la capacidad de aprendizaje de los humanos con pocos o cero ejemplos, ha habido un reciente surgimiento del interés en el aprendizaje automático reducido que utilice pocos ejemplos o cero ejemplos.

Los seres humanos pueden distinguir al menos 30,000 categorías de objetos básicos y muchas más subordinadas (por ejemplo, razas de perros). También pueden crear nuevas categorías de forma dinámica a partir de unos pocos ejemplos o simplemente basándose en una descripción de alto nivel. En contraste, la mayoría de las técnicas de visión por computadora existentes requieren cientos, si no miles, de muestras etiquetadas para cada clase de objeto con el fin de aprender un modelo de reconocimiento. Inspirado por la capacidad de los seres humanos para reconocer sin ver muchos ejemplos, el área de investigación de aprender a aprender o aprendizaje permanente [5,6] ha recibido un interés creciente.

**¿Es posible diseñar un modelo de aprendizaje automático con propiedades similares: aprender nuevos conceptos y habilidades rápidamente con solo algunos ejemplos de capacitación?**
**Eso es esencialmente lo que el aprendizaje simplificado pretende resolver.**

## Meta-aprendizaje

El éxito de las técnicas de aprendizaje profundo se ha limitado en gran medida a dominios en los que se dispone de abundantes datos y grandes recursos informáticos. La razón de esto es que aprender una nueva tarea requiere una gran cantidad de recursos. El meta-aprendizaje es un enfoque que promete relajar estos requisitos aprendiendo a aprender. Este campo ha atraído mucha atención en los últimos años. Un buen modelo de aprendizaje automático a menudo requiere ser "entrenado", es decir, expuesto a un gran número de muestras o ejemplos de entrada. Los humanos, en contraste, aprenden nuevos conceptos y habilidades mucho más rápido y más eficientemente. Los niños que han visto gatos y pájaros solo unas pocas veces pueden distinguirlos rápidamente. Las personas que saben cómo andar en bicicleta es probable que descubran la forma de conducir una motocicleta rápidamente con poca o ninguna demostración.

Se espera que un buen modelo de meta-aprendizaje sea capaz de adaptarse bien o generalizar nuevas tareas y nuevos entornos, aún cuando nunca se han mostrado durante el tiempo de entrenamiento. El proceso de adaptación, esencialmente una mini sesión de aprendizaje, ocurre durante la prueba pero con una exposición limitada a las nuevas configuraciones de tareas. Eventualmente, el modelo adaptado puede completar nuevas tareas. Es por esto que el meta-aprendizaje también se conoce como aprender a aprender.

Algunos ejemplos de tareas de meta-aprendizaje son:

1. Un clasificador entrenado en imágenes que no sean gatos puede decir si una imagen dada contiene un gato después de ver un puñado de imágenes de gatos.
2. Un bot de juego puede dominar rápidamente un juego nuevo.
3. Un mini robot completa la tarea deseada en una superficie cuesta arriba durante la prueba, incluso aunque solo se entrenó en un entorno de superficie plana.

## Modelos de aprendizaje automático simplificado

El diseño de un modelo de aprendizaje simplificado debe considerar dos características principales:

I Tener la capacidad de reconocer sin utilizar muchos ejemplos (sin una muestra, con una o algunas pocas muestras).

II Conservar el conocimiento aprendido de tareas anteriores y utilizarlo para ayudar al aprendizaje futuro (aprendizaje por transferencia)

Actualmente, los modelos de aprendizaje automático simplificado, se pueden dividir en dos grandes enfoques, aquellos que se basan en el *aprendizaje directamente supervisado* y aquellos basados en la *transferencia del aprendizaje*.

Los modelos de aprendizaje directamente supervisado, son aquellos que utilizan los ejemplos muestra para entrenar un clasificador estándar mediante aprendizaje supervisado (los ejemplos están bien definidos previamente) y se caracterizan porque:

- No utilizan datos auxiliares, (aquellos relacionados con los ejemplos muestra de dónde se extraen conocimientos transferibles para compensar la falta de ejemplos para entrenar el modelo)
- Aprende directamente por medio de un clasificador (los ejemplos muestra son seleccionadas como la única entrada para el modelo)

Por otra parte, los modelos basados en la transferencia del aprendizaje, pueden identificarse porque:

- Utilizan el conocimiento de los datos auxiliares
- Se basan en el paradigma de aprender a aprender o meta-aprendizaje.

Dentro de los modelos basados en la transferencia de aprendizaje, podemos dividir los algoritmo en tres principales categorías:

(a) Algoritmos basados en atributos

(b) Algoritmos de meta-aprendizaje

(c) Algoritmos de aprendizaje métrico

Los **algoritmos basados en atributos** utilizan el paradigma de aprender conjuntamente atributos latentes y definidos por el usuario, proporcionando un mecanismo intuitivo para acortar la brecha semántica y modelar etiquetas dispersas, incompletas y ambiguas. Un atributo, por ejemplo tiene alas, la cual es una característica intrínseca que posee el objeto o una clase (ejemplo, pájaro), y puede también indicar propiedades o anotaciones (por ejemplo, tiene una cabeza) de una imagen o un objeto. Los atributos describen una clase o una instancia, en contraste con la clasificación típica en la que se proporciona un nombre a esa clase. Un conjunto más rico de atributos pueden incluir partes, formas, materiales e incluso el reconocimiento de acciones humanas.

Como ejemplo de estos algoritmos se pude mencionar al *Modelo de Atributo Latente MultiModal* (M2LATAM, multi-modal latent attribute topic model), cuya aportación es que se centra en vídeos de actividades de grupos sociales. Los atributos de este algoritmo son, entre otros, el reconocimiento de acciones humanas, objetos y sonidos dentro de las imágenes. Estos atributos dentro de las muestras son particularmente desafiantes y de actualidad debido a su contenido multimodal y su naturaleza compleja y no estructurada en relación con la cantidad de información que pueda estar etiquetada y/o clasificada (ver Figura 2).

Los **algoritmos de Meta-aprendizaje** entrenan un modelo en una variedad de tareas de aprendizaje, de modo que pueda resolver nuevas tareas de aprendizaje utilizando sólo una pequeña cantidad de muestras o ejemplos de entrenamiento. En estos algoritmos, los parámetros del modelo se entrenan explícitamente de modo que, una pequeña cantidad de pasos acoplados con una pequeña cantidad de datos de entrenamiento para una nueva tarea, producirán un buen rendimiento de generalización en esa tarea. Estos algoritmos exploran el paradigma de aprender a aprender y su objetivo es transferir el conocimiento del conjunto de datos auxiliares al conjunto de datos de destino con uno o pocos ejemplos por objeto o clase. En concreto, el conocimiento se puede extraer y compartir en forma de características, atributos semánticos o información contextual.

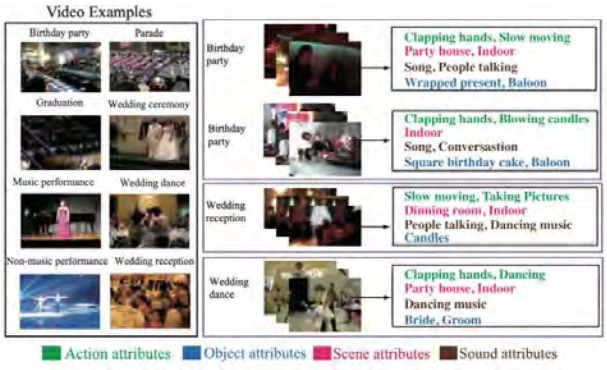



**Figura 2.** Ejemplo de clasificación del modelo M2LATAM (adaptado de [7]).

Como ejemplo de estos algoritmos se puede mencionar al *Modelo de Meta aprendizaje utilizando memoria a corto plazo* (META-LEARN LSTM, Meta Learn Long Short Term Memory) [7], cuyo funcionamiento se basa en utilizar un optimizador de meta-aprendizaje que está diseñado para optimizar el clasificador de una red neuronal artificial. El optimizador captura tanto el conocimiento a corto plazo dentro de una tarea como el conocimiento a largo plazo común entre todas las tareas. Mediante el uso de un objetivo, que captura directamente la capacidad del algoritmo para tener un buen rendimiento de generalización con solo un número determinado de actualizaciones, el modelo de meta-aprendizaje se entrena para hacer converger rápidamente el clasificador en una buena solución para cada tarea.

En la Figura 3 se presenta un ejemplo de la configuración de meta-aprendizaje que se utiliza en el algoritmo META-LEARN LSTM. La parte superior representa el conjunto de meta-entrenamiento (*Dmeta-train*), donde se muestra dentro de cada cuadro gris que existe un conjunto de datos separados, que a su vez consta del conjunto de entrenamiento *Dtrain* (lado izquierdo de la línea punteada) y el conjunto de prueba *Dtest* (lado derecho de la línea discontinua). En esta figura se considera la tarea de clasificación de 1 muestra y 5 clases, donde para cada conjunto de datos, tenemos un ejemplo de cada una de las 5 clases (a cada una se le da una etiqueta del 1 al 5) en el conjunto de entrenamiento y 2 ejemplos para evaluación en el conjunto de prueba. El conjunto de meta-prueba *Dmeta-test* se define de la misma manera, pero con un conjunto diferente de datos que cubren clases que no están presentes en ninguno de los conjuntos de datos en *Dmeta-train*.

Los **algoritmos de aprendizaje métrico** se centran en el aprendizaje transferible y predefinen una métrica fija (por ejemplo, como la distancia euclidiana), además tienen como objetivo aprender una métrica profunda transferible para comparar la relación entre imágenes (aprendizaje de pocas tomas), o entre imágenes y

descripciones de clases (aprendizaje sin tomas). De manera general, estos modelos realizan el reconocimiento con pocas tomas al aprender a comparar imágenes de consulta con algunas imágenes de muestra etiquetadas. Primero, utilizan un módulo integrado que genera representaciones de la consulta y de las imágenes de entrenamiento. Posteriormente, éstas se comparan mediante un módulo de relación que determina si son de categorías que coinciden o no.

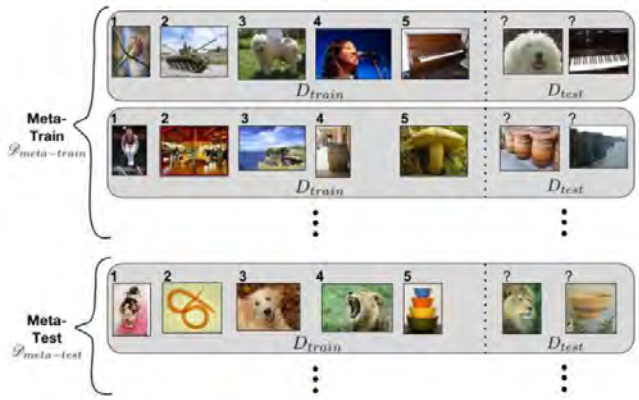



**Figura 3.** Ejemplo de configuración de meta-aprendizaje para el algoritmo META-LEARN LSTM (tomado de [7]).

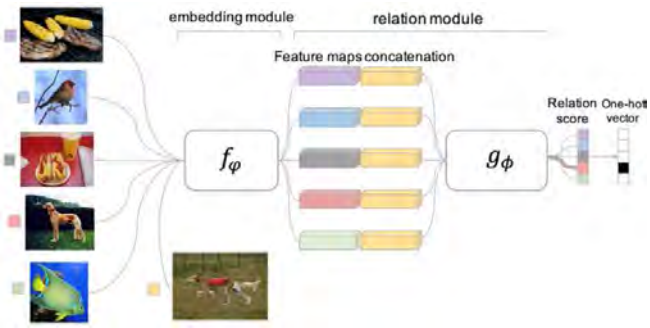



**Figura 4.** Arquitectura de una Red de relaciones (RELATION NET) para un problema de 5 muestras únicas con un ejemplo de consulta (tomado de [8]).

Como ejemplo de estos algoritmos se pude mencionar la denominada Red de Relaciones (RN, Relation Network), que durante el proceso de meta-aprendizaje, aprende una métrica de distancia para comparar una pequeña cantidad de imágenes dentro de cada episodio de entrenamiento, donde cada una está diseñada para simular la configuración de pocas muestras o tomas. Una vez entrenada la red de relaciones, puede clasificar imágenes de nuevas clases calculando relación entre las imágenes de consulta y los pocos ejemplos de cada nueva clase sin necesidad de actualizar más la red.

## Aplicaciones

Debido a que muchos de los posibles problemas a resolver, mediante aprendizaje automático, no cuentan con una base de muestras amplio para entrenamiento, así como el esfuerzo para la recopilación de estos datos y el costo computacional que implican los sistemas que utilizan muchas muestras, es que la utilización del aprendizaje automático simplificado ha surgido como una alternativa. Este tipo de aprendizaje a su vez, permite tener una forma más natural de poder imitar el aprendizaje similar al humano. Por lo tanto, muchas aplicaciones del mundo real podrán involucrar aprendizaje automático simplificado. La visión por computadora es uno de los primeros candidatos de pruebas para éste tipo de algoritmos. Esta forma de aprendizaje automático también ha atraído mucha atención recientemente en muchas otras aplicaciones, como la robótica, el procesamiento del lenguaje natural y el procesamiento de señales acústicas.

**Visión por computadora**. Las dos aplicaciones más populares son el reconocimiento de caracteres [8] y la clasificación de imágenes. Ya se han obtenido precisiones muy altas en los conjuntos de datos de referencia estándar (como Ominiglot y miniImageNet). Además del reconocimiento de caracteres y la clasificación de imágenes, también se han considerado otras aplicaciones de imágenes. Estos incluyen reconocimiento de objetos, transferencia de estilo de fuente, recuperación de imágenes, seguimiento de objetos, recuento de objetos específicos en imágenes, reconocimiento de ubicación de escenas, reconocimiento de gestos, etiquetado de partes, generación de imágenes y reconstrucción de formas para objetos 3D. También se ha utilizado con éxito en aplicaciones de vídeo, incluida la predicción de movimiento, la clasificación de vídeo, la localización de acciones, la re-identificación de personas, la detección de eventos y la segmentación de objetos.

**Robótica**. Para que los robots se comporten más como humanos, deben poder generalizar su conocimiento y aprendizaje, a partir de poder observar y entender algunas demostraciones o ejemplos. Por tanto, el aprendizaje automático simplificado jugará un papel importante en la robótica. Por ejemplo, aprender el movimiento del brazo de un robot usando el aprendizaje de imitación de una sola demostración, y aprender acciones de manipulación de algunas demostraciones con la ayuda de un maestro que corrige las acciones falsas [9]. Además de imitar a los usuarios, los robots pueden mejorar su comportamiento interactuando con los usuarios. Recientemente, las estrategias de asistencia se aprenden a partir de algunas interacciones a través del aprendizaje por refuerzo de utilizando aprendizaje automático simplificado. Otros ejemplos incluyen navegación visual y control continuo en entornos dinámicos.

**Procesamiento de lenguaje natural**. Las aplicaciones incluyen análisis sintáctico, traducción, finaliza-

ción de oraciones (que completa los espacios en blanco con una palabra elegida de un conjunto proporcionado), clasificación de opiniones de reseñas breves, clasificación de la intención del usuario para los sistemas de diálogo, la predicción de cargos penales y la clasificación de texto de múltiples etiquetas.

**Procesamiento de señales acústicas**. Los esfuerzos recientes se centran en la síntesis de voz . Una tarea popular es la clonación de voz a partir de algunas muestras de audio del usuario. Esto puede ser útil para generar navegación por voz personal en aplicaciones de mapas o imitar la voz de los padres al contar historias a los niños en un juego de herramientas para el hogar inteligente. Recientemente, es posible realizar la conversión de voz de un usuario a otro utilizando una sola muestra de voz o texto e incluso en diferentes idiomas [10].

En resumen, existen muchos campos y aplicaciones interesantes que el aprendizaje automático simplificado puede explorar.

## Retos actuales del aprendizaje automático simplificado

Aunque los últimos años ha existido un progreso tanto en la metodología como en las aplicaciones del aprendizaje simplificado, aún existen desafíos debido a la dificultad intrínseca de los ejemplos o muestras escasas. En esta sección, se sugieren tres retos a considerar.

**Robustez**. La mayoría de los estudios y modelos actuales sobre el aprendizaje automático simplificado, se basan en la hipótesis de que se utilizarán datos ideales, pero es difícil mantenerlo cierto en todas las escenas prácticas. En muchas aplicaciones realistas, uno puede enfrentarse a perturbaciones inciertas que destruyen el escenario ideal. Por ejemplo, los pocos datos de entrenamiento pueden sufrir interferencias atípicas, por ejemplo, imágenes borrosas debido a un mal funcionamiento de la cámara que tomó dicha imagen. Por lo tanto, mejorar la robustez de los modelos frente a varios factores de perturbación potenciales es sustancialmente significativo.

**Universalidad**. La universalidad mencionada se refiere, por una parte, a la generalidad y escalabilidad a nivel de modelo de los enfoques del aprendizaje automático simplificado. Actualmente, la mayoría de los modelos están excesivamente diseñados para las tareas y conjuntos de datos de referencia específicos, lo que debilita su aplicabilidad a otras tareas más generales. Un modelo ideal debería poder manejar varias tareas de aprendizaje con diferente complejidad de datos y diversas formas de datos. Por otro lado, la universalidad también se refiere a la versatilidad y flexibilidad a nivel de aplicación de los enfoques de éste tipo de aprendizaje. La mayoría de los estudios actuales se centran en el escenario de aplicación simple con clases de tareas a pequeña escala, sin embargo, los problemas del mundo real pueden traer escenarios de aplicación más complejos, como clases de tareas a gran escala, falta de disponibilidad de datos etiquetados correctamente e incluso una combinación de estos escenarios.

**Interpretabilidad**. El auge y el éxito de los modelos de aprendizaje automático en los últimos años radican principalmente en la tecnología de aprendizaje profundo, implementadas en redes neuronales de varias capas, que a menudo son cuestionadas por su falta de interpretabilidad. La interpretabilidad del modelo es una cuestión clave para el aprendizaje profundo, es decir, que los resultados obtenidos puedan explicarse y predecirse basados en el análisis del comportamiento de la red neuronal y a su vez este conocimiento previo pueda utilizarse nuevamente en el modelo. Por lo tanto, el poder capitalizar la fusión de conocimiento previo externo y conocimiento de datos internos para mejorar la interpretabilidad de los modelos es un gran reto y podría ser una dirección de investigación futura.

## Conclusiones

Los objetivos de aprendizaje automático reducido consisten en cerrar la brecha entre la llamada Inteligencia Artificial y el aprendizaje humano.

Un sistema artificial utilizando éste modo de aprendizaje, puede aprender nuevas tareas que contienen solo algunos ejemplos con información supervisada incorporando conocimientos previos. Se puede decir que actúa como un banco de pruebas para la Inteligencia Artificial, haciendo posible el aprendizaje de casos aislados o excepcionales, que puedan caracterizarse a partir del aprendizaje previo, así mismo, ayuda a aliviar la carga de recopilar datos supervisados a gran escala para diversas aplicaciones.

El uso exitoso de algoritmos de meta-aprendizaje en escenarios donde ocurren cambios en la distribución de tareas sigue siendo un importante desafío abierto en el campo del meta-aprendizaje. Por esta razón, el avance en el diseño de algoritmos en éste sentido, pueden ser una pieza importante del rompecabezas para ampliar la aplicabilidad de las técnicas de meta-aprendizaje a problemas del mundo real.

Reutilizar el conocimiento de tareas previas puede ser un ingrediente crucial para hacer modelos escalables de alta capacidad, como redes neuronales profundas, susceptibles de entrenamiento rápido con pequeños conjuntos de datos. Esto sería un paso hacia una técnica de meta-aprendizaje simple y de propósito general que se puede aplicar a cualquier problema y modelo. La investigación adicional en esta área puede hacer que el aprendizaje automático simplificado sea un ingrediente estándar en el aprendizaje profundo y el aprendizaje por refuerzo. ✳

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

## REFERENCIAS


1. Harikiran, G., Menasinkai, K. y Shirol, S. (2016). Smart security solution for women based on Internet Of Things (IOT). En *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*. Chennai, India. 3551-3554.

2. Timmons, A., Chaspari, T., Han, S., Perrone, L., Narayanan, S. y Margolin, G. (2017). Using multimodal wearable technology to detect conflict among couples. *Computer*, 50(3), 50-59.

3. Giannakopoulos, T., Kosmopoulos, D., Aristidou, A. y Theodoridis, S. (2006). Violence content classification using audio features. En *Hellenic Conference on Artificial Intelligence*. Berlin, Heidelberg, 502-507.

4. Lee, Y., Kim, K., Han, D. y Ko, H. (2012). Acoustic and visual signal based violence detection system for indoor security application. En *IEEE International Conference on Consumer Electronics (ICCE)*. Las Vegas Nv, USA. 737-738.

5. Bautista-Duran, M., García-Gomez, J., Gil-Pita, R., Sánchez-Hevia, H., Mohino-Herranz, I.y Rosa-Zurera, M. (2017). Acoustic Detection of Violence in Real and Fictional Environments. En *ICPRAM*. 456-462.

6. Andersson, M., Ntalampiras, S, Ganchev, T., Rydell, J., Ahlberg, J., Division of Information Systems, FOI Swedish Defence Research Agency, Sweden, L. y Fakotakis, N. (2010). Fusion of acoustic and optical sensor data for automatic fight detection in urban environments. En *13th International Conference on Information Fusion*. Edinburg, UK. 1-8.

7. Grega, M., Matiolański, A., Guzik, P. y Leszczuk, M. (2016). Automated detection of firearms and knives in a CCTV image. *Sensors*, 16(1), 47.

8. Shulby, C., Pombal, L., Jordao, V., Ziolle, G., Martho, B., Postal, A. y Prochnow, T. (2018). Proactive Security: Embedded AI Solution for Violent and Abusive Speech Recognition. En *7th Brazilian Conference on Intelligent Systems (BRACIS)*. Sao Paulo, Brazil. 248-253.

9. Subramani, S., Vu, S. y Wang, H. (2017). Intent classification using feature sets for domestic violence discourse on social media. En *4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*. Mana Island, Fiji. 129-136.

10. Just, M., Chandok, H., Sampangi, R., Hawkey, K., Willis, A., Samuthiravelu, J., Gill, D. y Altair, M. (2019). Personal safety app effectiveness. En *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow Scotland, UK. 1-6.


## SOBRE LOS AUTORES

**Alejandro Hernández-Herrera** es estudiante de doctorado en el Laboratorio de Robótica y Mecatrónica en el Centro de Investigación en Computación (CIC) del Instituto Politécnico Nacional (IPN). Obtuvo su grado de Maestría en Ciencias en Ingeniería de Cómputo en el Centro de Investigación en Computación del Instituto Politécnico Nacional en 2018. Sus intereses científicos incluyen sistemas inmersivos, sistemas bio-inspirados, robótica de enjambre y aprendizaje automático.

**Elsa Rubio Espino** obtuvo el grado de Doctora en Ciencias en la Especialidad de Sistemas Automáticos por el Instituto Nacional de Ciencias Aplicadas de Toulouse, Francia. Actualmente es profesora-investigadora del Centro de Investigación en Computación (CIC) del Instituto Politécnico Nacional (IPN), adscrita al Laboratorio de Robótica y Mecatrónica. Sus intereses científicos incluyen la robótica móvil y de servicio, así como la dinámica de sistemas a eventos discretos.

**Rogelio Álvarez Vargas** es Ingeniero en Comunicaciones y Electrónica de la Escuela Superior de Ingeniería Mecánica y Eléctrica (ESIME) del Instituto Politécnico Nacional (IPN) en 1983, realizó la Maestría en Ciencias Ingeniería Eléctrica con especialidad en control automático en el Centro de Investigación y de Estudios Avanzados (CINVESTAV) en 1985, y el Doctorado en el Instituto politécnico de Grenoble, Francia; especialidad Automática y Tratamiento de Señal en 1992. Actualmente es Director de TI, Electrónica y Control en el Centro de Tecnología Avanzada (CIATEQ, A. C.). Sus intereses de investigación son los sistemas de manufactura inteligente y la robótica de servicio.

**Víctor Hugo Ponce Ponce** obtuvo el grado de Doctor en Ciencias en la Especialidad de Ingeniería Eléctrica en el Centro de Investigación y de Estudios Avanzados (CINVESTAV) en 2005. Actualmente es profesor-investigador del Centro de Investigación en Computación (CIC) del Instituto Politécnico Nacional (IPN) y Jefe del Laboratorio de Microtecnología y Sistemas Embebidos del CIC. Sus intereses científicos incluyen los sistemas neuromórficos, el diseño de circuitos integrados y los MEMS.

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

ARTÍCULO ACEPTADO

# SASTuit, software de análisis de sentimiento utilizando aprendizaje automático

Anibal A. Herrera-Contreras, Eddy Sánchez-DelaCruz, Ivan V. Meza-Ruiz y Mirta Fuentes-Ramos

## El impacto de las redes sociales en la obtención de datos para el análisis de sentimiento

Hoy en día, con la creciente popularidad de las redes sociales, nos encontramos ante una gran variedad de medios digitales que permiten compartir con facilidad opiniones, experiencias, posturas y sentimientos sobre diversos temas de nuestro interés. *Twitter* es una de las plataformas más utilizadas en el mundo [1]; en ella se generan alrededor de 250 millones de *tweets* diarios, convirtiéndose en el medio idóneo para la aplicación de la minería de opiniones y el análisis de sentimiento.

De acuerdo a Social, W. A. (2020) existen 4 540 millones de internautas en el mundo, mientras que en las redes sociales hay 3 800 millones de usuarios [1]. Son cada vez más las personas que incorporan estas aplicaciones modernas en su vida cotidiana; como consecuencia ha surgido una demanda en el desarrollo de nuevas tecnologías emergentes capaces de manipular y analizar grandes volúmenes de datos, obteniendo de ellos, patrones, tendencias y conclusiones que ayuden a entender de mejor manera a los usuarios y así tomar decisiones en rubros comerciales, electorales o en su caso generar estrategias de *marketing* exitosas sobre productos o servicios [2, 3, 4].

## Antecedentes

El análisis de sentimiento es un tipo de procesamiento del lenguaje natural (PLN) que permite conocer el estado de ánimo de los usuarios sobre algún tema en particular; es un término que se encuentra muy ligado con las redes sociales, ya que es allí donde la mayoría de los usuarios expresan su opinión; no obstante no se limita a ellas.

Liu, B. (2012) lo define como el campo de estudio encargado de analizar las opiniones, los sentimientos y emociones de las personas hacia entidades como productos/servicios, organizaciones y temas con el objetivo de conocer la actitud del escritor [5].

Identificar el sentimiento predominante de los usuarios es una tarea muy compleja, incluso para el ser humano y es la razón de ser de esta disciplina. En los últimos años se ha realizado un enorme número de estudios de análisis de sentimiento y se ha aplicado en una gran variedad de campos interdisciplinarios, como son la política, la tecnología, la medicina, el sector empresarial, por mencionar algunos. Cabe destacar que la mayoría de los estudios en esta rama utilizan la red social *Twitter* como fuente principal de obtención de datos para analizar la opinión pública.

Por lo anterior, se ha vuelto una necesidad en los corporativos el monitoreo de las redes sociales, con la finalidad de analizar las opiniones de sus clientes y obtener una retroalimentación sobre sus productos/servicios, esto para conseguir una mejora continua, pues este tipo de estudios permite a las empresas realizar investigaciones de mercado sin la necesidad de recurrir a encuestas en personas, obteniendo así una mayor información de calidad.

Con el paso del tiempo, el análisis de sentimiento se ha convertido en un tema de investigación jerárquico debido a la gran demanda existente en el mercado por analizar la opinión pública. Sus orígenes se remontan al año de 1996, con el desarrollo del sistema *General Inquirer*, el cual agrupaba las palabras en diversas categorías [6].

Hoy por hoy, existen diversas técnicas, métodos, bibliotecas y herramientas para llevar a cabo la implementación del PLN, con la finalidad de obtener conocimiento valioso en los datos que puedan ayudar a solventar alguna necesidad social. Por ejemplo, en una investigación realizada por Saura *et al.* (2018), se muestra un análisis de sentimientos basado en el *hashtag #BlackFriday* utilizando aprendizaje automático. Su objetivo principal fue identificar la identidad del sentimiento de los consumidores en relación con las tres clases *{positivos, negativos, neutrales}*. Ellos intentaban vincular este sentimiento con las ofertas publicadas por diferentes compañías. Usaron la herramienta en línea *Monkeylearn*; como resultado el 32 % de los *tweets* fueron identificados como positivos, 7.7 % como negativo y 60.2 % como neutral. Sus resultados apuntan a producir mejores estrategias de *marketing* [7].

Por otro lado, Martin-Domingo *et al* (2019) analizaron opiniones sobre la cuenta de Londres Heathrow en *Twitter* desde el punto de vista del análisis de sentimientos para identificar nuevas ideas que podrían mejorar la calidad del servicio del aeropuerto (ASQ, por las siglas en inglés de *airport service quality*). Desarrollaron un sistema que podría recopilar y analizar los datos. Su diseño incluyó una lista de 108 palabras clave para la identificación de *tweets* que hablan sobre la calidad de los *tweets* que se ocupan de la ASQ. Su trabajo concluyó que había 23 atributos sobre el ASQ. El análisis de sentimientos se realizó utilizando las herramientas *Theysay* y *Twin-*

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

*word*, cuyos resultados fueron verificados manualmente. Dichos resultados mostraron una precisión de 78.7 % y 69.6 % respectivamente [8].

Joyce y Deng (2017) realizaron un análisis de sentimiento en *Twitter* para las elecciones presidenciales de EE. UU. del 2016. Para ello recopilaron *tweets* que tuvieran relación con los nombres de los candidatos días antes de las elecciones, después calcularon el sentimiento empleando dos enfoques: análisis de sentimiento basado en el léxico utilizando *OpinionFinder* y análisis de sentimiento con aprendizaje automático haciendo uso de la herramienta *NLTK* para implementar el algoritmo *Naive Bayes*; luego compararon los resultados obtenidos con encuestas para identificar la correlación existente que comparten. El estudio demostró que existe un coeficiente de correlación muy alto del 94 % con los datos de las encuestas; además mencionan que quizás en el futuro, las encuestas en las redes sociales se incorporen más a los esquemas de votación [9].

Para conocer más detalles acerca del estado-del-arte del PLN y el Análisis de Sentimiento, se recomienda consultar [10, 2]. Los resultados que ha mostrado el análisis de sentimiento, en conjunto con el aprendizaje automático aplicado en diversas áreas interdisciplinarias, son realmente prometedores.

**El Procesamiento de Lenguaje Natural es una tendencia creciente con amplio futuro y diversas áreas de aplicación.**

En los trabajos descritos anteriormente los autores hicieron uso de varias herramientas y bibliotecas para poder llevar a cabo con éxito este tipo de estudios; si bien cada una de ellas han ayudado para la implementación de la minería de opinión y análisis de sentimiento, ninguna aborda de manera general la solución; es decir, un *software* que pueda realizar tareas de obtención de datos, limpieza de datos, predicción y visualización de resultados.

En la Figura 1 se muestra de manera general la metodología que utilizan diversos autores para llevar a cabo la implementación de Análisis de Sentimiento, abordando cuatro fases principales.

**Figura 1.** Metodología general del análisis de sentimiento.

- Fase 1 - Recolección de datos: Esta fase implica la búsqueda y recopilación de datos; si bien no existe un estándar para llevar a cabo este proceso, la mayoría de los trabajos publicados toman a *Twitter* como fuente rica para obtención de datos haciendo uso de su *API* para conectarse y acceder a datos sobre los usuarios para su análisis.

- Fase 2 - Limpieza de datos: La segunda fase consiste en la limpieza y preparación de los datos. En este punto se eliminan aquellas palabras que no contienen mucha información sobre el sentimiento en el texto, comúnmente conocidas como *stopwords*. Normalmente suelen estar conformadas por pronombres, preposiciones, adverbios y verbos, así como también se eliminan datos duplicados como *retweets*, páginas *web*, correos electrónicos, *hashtags*, etiquetas de usuarios, caracteres especiales como signos de puntuación, acentos, espacios en blanco y números, además de la conversión del texto de mayúsculas a minúsculas.

- Fase 3 - Detección del sentimiento: Para la detección del sentimiento existen dos vertientes principales: i) enfoque basado en el léxico y ii) enfoques basados en aprendizaje supervisado. El primer enfoque es conocido como la aproximación basada en orientación semántica; cada palabra definida en el diccionario está asociada a un valor numérico que define el sentimiento. Por ejemplo *SentiWordNet* es un léxico que asigna tres puntuaciones de sentimiento tales como positividad, negatividad u objetividad a un texto dado; por otro lado, se encuentra el léxico de opiniones de Bing Liu que determina la polaridad del texto en dos clases *{positivos, negativos}*. Este léxico se creó con la recopilación de datos obtenidos desde diferentes sitios de la *web* creando un léxico informal puesto que normalmente no siempre se siguen las reglas gramaticales en sitios de internet.

- Fase 4 - Validación de resultados: Por último, se implementan diversas métricas de validación, como la matriz de confusión, sensibilidad, especificidad, curva ROC, entre otras, para evaluar la confiabilidad de un modelo de aprendizaje automático, en términos de exactitud, al momento de clasificar los datos.

## Herramienta SASTuit

Dado que el Análisis de Sentimiento es una de las áreas de investigación que está en auge debido a las bondades como método para conocer tendencias de los usuarios respecto a algún tema en específico, se desarrolló un

modelo general para adquirir información de las redes sociales (específicamente *twitter*), pre-procesarla y clasificar de acuerdo a las emociones de casos de estudios variados [2].

**La opinión pública es un fenómeno social que puede aludir al éxito si es atendida en su momento.**

SASTuit es un *software* capaz de analizar el comportamiento de los usuarios ante cualquier tema de interés en las redes sociales utilizando un enfoque de aprendizaje supervisado para la clasificación del sentimiento predominante y conocer de mejor manera la opinión pública.

Es una herramienta versátil para llevar acabo la implementación de análisis de sentimiento de forma automatizada; consta de cuatro funciones principales, tales como 1) búsqueda de *tweets*, 2) la unión de varios conjuntos de datos, 3) tareas de pre-procesamiento y filtrado de datos y 4) la predicción de *tweets* en dos clases *{positivos, negativos}* en español.

El proceso metodológico implementado se muestra en la Figura 2 y se divide en dos fases: i) la fase experimental o de entrenamiento y ii) la fase de desarrollo del software.

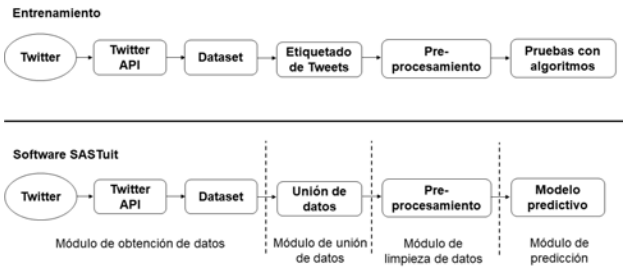



**Figura 2.** Proceso metodológico.

Fase 1: Entrenamiento.
La primera fase implica la recopilación de datos de los usuarios de *Twitter* usando *RStudio* como lenguaje de programación y la biblioteca *Rtweet* para conectarse a la *API* de *Twitter* etiquetando manualmente cada *tweet* en dos clases tales como *{positivos, negativos}* en idioma español.

- Obtención de datos: Se utilizó la *API* de búsqueda de *Twitter* buscando *#hashtags* y temas en tendencia en esta red social; también se consideraron las cuentas de usuarios referentes a noticias, productos/servicios, librerías y figuras públicas. Asimismo se consideraron ciertas palabras clave tales como amor, odio, feliz, excelente servicio, entre otras, con la finalidad de tener un conjunto de datos rico en temas variados.

- Etiquetado manual de *tweets*: La muestra de datos para esta investigación está conformada por 2,282 *tweets* (un valor intermedio en relación a otros trabajos descritos en el estado-del-arte), con 964 positivos y 1,318 negativos; se supervisó cada *tweet* seleccionando y etiquetando sólo aquellos que contienen mayor información sobre la polaridad del sentimiento. También se corrigieron palabras y abreviaturas mal escritas como por ejemplo fav/favorita, amooor/amor, app/aplicación, entre otras.

- Pre-procesamiento y preparación de datos: Como parte del pre-procesamiento de datos a la muestra final se normalizaron los datos convirtiendo el texto a minúsculas; también se eliminaron aquellas palabras que no contienen mucha información para identificar el sentimiento pero sin excluir las palabras «no» y «si».

- Prueba con algoritmos de aprendizaje supervisado: Mediante validación cruzada, se implementaron seis algoritmos supervisados como métodos para la realización de pruebas en la clasificación de sentimientos: bosques aleatorios, regresión logística, naive bayes, potenciación del gradiente, árboles de decisión y máquinas de vectores de soporte; esto para definir el clasificador con mayor precisión en los datos. Se usaron dichos clasificadores debido a su simplicidad y efectividad en la predicción del sentimiento, además de ser los más utilizados en otras investigaciones.

  Para la validación de resultados se realizaron experimentos en el conjunto de datos etiquetados para la creación de un modelo predictivo de sentimientos a partir de datos conocidos y se calculó la precisión con datos nuevos. Un estándar que la mayoría de trabajos publicados utilizan es la regla 80-20, dónde el 80 % de los datos son para entrenamiento y 20 % para validación.

  El algoritmo que obtuvo un mejor rendimiento durante la fase experimental fue Naive Bayes con 92 % en precisión, sensibilidad, puntaje $f_1$ y un 91.90 % de exactitud en la clasificación del sentimiento.

Fase 2. Software SASTuit.
La segunda fase consistió en automatizar la fase 1, utilizando la biblioteca *Shiny* de *R*, obteniendo una interfaz gráfica de usuario interactiva y amigable al usuario final.

- Módulo de obtención de datos: Este módulo tiene como objetivo automatizar el proceso de recolección de datos de las publicaciones compartidas por los usuarios en *Twitter* realizando búsquedas por medio de *hashtags*, palabras clave o nombres de usuarios. Para iniciar el proceso de obtención de datos desde la herramienta SASTuit es necesario conectarse a la *API* de *Twitter* ingresando los *tokens* de acceso.

- Módulo de unión de datos: El segundo módulo tiene como objetivo la unión de varios archivos con extensión CSV; esto con el fin para homologar en un solo documento todos los *tweets* o textos que se desean analizar. La herramienta toma en cuenta los nombres de las columnas que tienen en común para la combinación de archivos, considerando que en el módulo anterior, con el tiempo, se pueden descargar varios conjuntos de datos a partir del parámetro *status_id*.

- Módulo de limpieza de datos: Este módulo tiene como objetivo automatizar el proceso de limpieza de datos; en esta fase de la metodología de análisis de sentimientos destaca la eliminación de ruido en el conjunto de datos, filtrando aquella información en el texto que no ayuda en la detección del sentimiento.

- Módulo de predicción: El módulo tiene como objetivo clasificar los *tweets* en dos clases *{positivos, negativos}*, utilizando el mejor algoritmo en la clasificación de sentimientos previamente entrenado: Naive Bayes, el cual obtuvo el mejor rendimiento en la predicción de datos en comparación a los otros cinco algoritmos. En la Figura 3 se puede observar el módulo de predicción de la herramienta SASTuit.

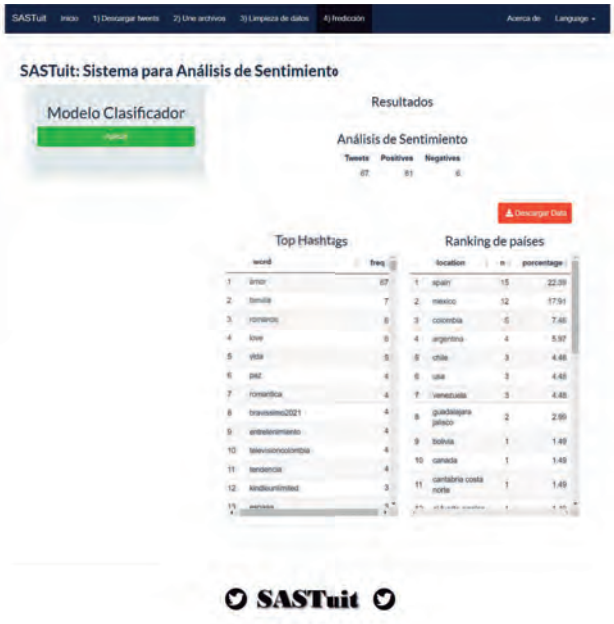



**Figura 3.** Herramienta SASTuit.

**Se desarrolló una herramienta libre (software on-line) capaz de recopilar y analizar grandes volúmenes de datos para la implementación del análisis de sentimientos y la minería de opiniones.**

## Limitaciones

SASTuit constituye una herramienta de fácil acceso y manipulación, permitiendo que el análisis de sentimiento esté al alcance de cualquier usuario; sin embargo, cuenta con un par de limitaciones, tales como:

1. La herramienta SASTuit sólo tiene soporte para analizar *tweets* en español.

2. La herramienta identifica el sentimiento en dos clases *{positivos, negativos}*, más no muestra la intensidad del sentimiento expresado.

## Aportaciones

SASTuit provee de oportunidades al sector empresarial para analizar las opiniones de sus clientes y medir la satisfacción de sus consumidores en términos de aceptación o rechazo sobre sus productos/servicios, permitiendo crear estudios de mercado y tendencias de una manera rápida, accesible y sin costo alguno. De igual forma se puede obtener un panorama general sobre la reputación de una marca y adquirir retroalimentación para la mejora continua; de esta manera se puede reducir la tasa de abandono de clientes, atendiendo sus inquietudes, necesidades y/o expectativas mejorando a su vez la percepción del usuario final.

El desarrollo de este proyecto aporta a la comunidad científica una nueva alternativa para la realización de análisis de sentimiento, ya que se deja a disposición gratuita el *corpus* de redes sociales (*twitter*) con anotaciones humanas en dos clases *{positivos, negativos}* en idioma español para futuras investigaciones. Tanto SASTuit como el *corpus* se encuentran disponible para descarga desde `https://analisis-sentimiento.shinyapps.io/sastuit/`.

Otro de los beneficios de la herramienta SASTuit es que se encuentra automatizado; es decir, no se necesitan conocimientos previos para poder analizar algún tema de interés, por lo que cualquier persona del público en general podrá utilizarlo.

No menos importante y destacable es que SASTuit es un software gratuito y de código abierto, por lo que ofrece al usuario final la posibilidad de estudiar el código fuente del programa y adaptarlo según sus necesidades, así como también de hacer y de distribuir copias del mismo aportando nuevas ideas y métodos que mejoren su funcionamiento.

## Conclusiones

Las redes sociales se han convertido en uno de los principales medios de comunicación usados por millones de personas alrededor del mundo. *Twitter* por su parte es una de las plataformas más populares que permite a las personas expresar libremente su opinión ante cualquier tema de interés; es por ello que constituye una gran fuente de información para la minería de opiniones y el análisis de sentimiento.

En este trabajo se presentó una herramienta que permite la realización de análisis de sentimiento a partir de datos obtenidos de la red social *Twitter*, con funcionalidades como lo es la obtención de datos de forma automatizada. También permite la unión de varios conjuntos de datos para su respectivo análisis al igual que aplicar tareas de limpieza y filtrado de datos, clasificando cada tweet como *{positivo, negativo}* comparando seis métodos de clasificación tradicionales, seleccionando el algoritmo más preciso para la implementación en la clasificación de sentimientos en la herramienta SASTuit.

La automatización y uso de SASTuit permitió observar que:

- *Twitter* constituye una fuente atractiva para la realización de análisis de sentimiento y la minería de opiniones.
- La clasificación de sentimiento de los *tweets* puede mejorar considerablemente si los datos se encuentran balanceados con respecto a las clases objetivo.
- La precisión en la clasificación del sentimiento depende en gran medida de la cantidad de datos con los que fueron entrenados los algoritmos.

Dado lo anterior, resulta evidente la importancia de desarrollar herramientas capaces de evaluar el análisis de sentimiento, para la obtención de conocimiento valioso y aplicarlo para la mejora ante diversas entidades en beneficio a la sociedad. ✷

## REFERENCIAS


1. Social, W.A. (2020). Global Digital Report 2020. Recuperado el 01 de Marzo de 2020, de `https://wearesocial.com/digital-2020`.
2. Herrera-Contreras, A.A., Sánchez-Delacruz, E., Meza-Ruiz, I.V. (2020). Twitter Opinion Analysis About Topic 5g Technology. En *Botto-Tobar M., Zambrano Vizuete M., Torres-Carrión P., Montes León S., Pizarro Vásquez G., Durakovic B. (eds) Applied Technologies. ICAT 2019. Communications in Computer and Information Science, Springer, Cham*, 1193, 191–203.
3. Jiménez-Márquez, J., González-Carrasco, I., López Cuadrado, J., y Ruiz-Mezcua, B. (2019). Towards a Big Data Framework for Analyzing Social Media Content. En *International Journal of Information Management*, 44, 1–12.
4. Sagiroglu, S., y Sinanc, D. (2013). Big data: A Review. En *2013 International Conference on Collaboration Technologies and Systems (CTS), IEEE*, 42–47.
5. Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5, 1–167.
6. Stone, P., Dunphy, D., y Smith, M. (1966). The General Inquirer: A Computer Approach to Content Analysis. *M.I.T. Press.*
7. Saura, J., Reyes-Menéndez, A., y Palos-Sánchez, P. (2018). Un Análisis de Sentimiento en Twitter con Machine Learning: Identificando el Sentimiento sobre las Ofertas de #BlackFriday. *Revista Espacios*, 39, 1–11.
8. Martin-Domingo, L., Martín, J., y Mandsberg, G. (2019). Social Media as a Resource Forsentiment Analysis of Airport Service Quality (ASQ). *Journal of Air Transport Management*, 78, 106–115.
9. Joyce, B., y Deng, J. (2017). Sentiment Analysis of Tweets for the 2016 us Presidential Election. En *2017 IEEE Mit Undergraduate Research Technology Conference (URTC)*, 1–4.
10. Herrera-Contreras, A. (2020). Modelo para Análisis de Sentimiento Utilizando Aprendizaje Automático. Tesis de Maestría, Maestría en Sistemas Computacionales. Instituto Tecnológico Superior de Misantla.


## SOBRE LOS AUTORES

**Anibal A. Herrera-Contreras** Es Ingeniero en Informática por el Instituto Tecnológico Superior de Perote, Veracruz, México y Maestro en Sistemas Computacionales por el Instituto Tecnológico Superior de Misantla, Veracruz, México. Actualmente labora como "Desarrollador Big Data y Data Science" en el sector privado. Sus intereses de investigación son el procesamiento de lenguaje natural y la minería de datos.

**Eddy Sánchez-DelaCruz** Es Doctor en Sistemas Computacionales por la Universidad del Sur, Maestro en Sistemas Computacionales y Licenciado en Informática Administrativa por la Universidad Autónoma de Tabasco. Actualmente es Profesor-Investigador en el Departamento de Posgrado e Investigación del Instituto Tecnológico Superior de Misantla, Veracruz, México. El Dr. Sánchez-DelaCruz es miembro del Sistema Nacional de Investigadores, nivel I, miembro de la RedICA-CONACyT, Perfil deseable PRODEP y fundador y líder del Cuerpo Académico Inteligencia Artificial. Sus intereses de investigación son el aprendizaje automático y la ciencia de datos con un enfoque multidisciplinario (`www.eddsac.com`).

**Ivan V. Meza-Ruiz** Es Ingeniero por la UNAM, México, Maestro y Doctor por la Universidad de Edimburgo. Investigador asociado al Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas de la UNAM. Especializado en la intersección entre el lenguaje humano y el cómputo usando técnicas de inteligencia artificial y aprendizaje automático. Ha participado en dieciséis proyectos de investigación en Inteligencia Artificial, autor de más de 80 artículos en congresos, talleres y revistas internacionales. Miembro del Sistema de Investigación Nacional nivel I.

**Mirta Fuentes-Ramos** Es Licenciada en Enfermería por el Instituto de Ciencias y Estudios Superiores de Tamaulipas y es Técnica en Enfermería por el Centro de Estudios Tecnológicos en Ciencias de la Salud, Actualmente es asistente de investigación en el Departamento de Posgrado e Investigación de la División Académica Multidisciplinaria de Comalcalco, Universidad Autónoma de Tabasco, México. Sus intereses de investigación son el cómputo Inteligente aplicado al área de la salud e ingeniería.

## *Minería de opiniones.*

La minería de opiniones, también conocida como ***análisis de sentimientos***, se refiere al uso del procesamiento del lenguaje natural, el análisis de texto y la lingüística computacional para identificar y extraer información subjetiva en los materiales de origen. El análisis de sentimientos se aplica ampliamente a las reseñas y las redes sociales para una variedad de aplicaciones, que van desde el marketing hasta el servicio al cliente. El objetivo del análisis de sentimientos es determinar la actitud de un hablante o un escritor con respecto a algún tema o la polaridad contextual general de un documento. La actitud puede ser su juicio o evaluación, el estado afectivo, es decir, el estado emocional del autor al escribir, o la comunicación emocional prevista, es decir, el efecto emocional que el autor desea tener en el lector.

*Imagen tomada de:* `https://coverimages.igi-global.com/images/dictionary-lp.png`.

ARTÍCULO ACEPTADO

# Panorama de la recuperación de música basada en estados emocionales

Mónica Muñoz-Ramírez, Humberto Pérez-Espinosa, José Alberto Fernández-Zepeda y Juan Martínez-Miranda

## Introducción

En este artículo se da un panorama de los avances actuales en el área de recuperación de música basada en el análisis acústico del audio de acuerdo a términos afectivos especificados por el usuario. Este tipo de tecnología se ha venido desarrollando en los últimos años provocado por el aumento gigantesco en la disponibilidad de música en línea. Lo anterior incrementa la necesidad de crear métodos que permitan organizar y recuperar la música deseada de manera eficiente y adecuada a los deseos del usuario. Las plataformas de música tradicionalmente asocian metadatos a las canciones de forma manual. Este proceso es costoso en tiempo y dinero ya que constantemente se están generando grandes cantidades de nuevas canciones. Los investigadores en el área de recuperación musical buscan crear métodos para organizar las canciones de acuerdo al estado de ánimo que expresan las mismas, utilizando descriptores acústicos y algoritmos de aprendizaje automático.

En las siguientes secciones se presenta la evolución de la forma de distribución de la música grabada y la problemática que representa el encontrar la música de nuestro agrado entre la inmensidad de música disponible actualmente. Se introducen las bases teóricas de los modelos de categorización emocional y las tendencias en los métodos de recuperación de música basada en aspectos emocionales.

## Necesidad de nuevas formas de recuperación de música

En los últimos años han habido cambios trascendentales en el formato de distribución y en la cantidad de música disponible. Esta evolución ha cambiado radicalmente la industria de la música, en la que la distribución a través de material físico ha quedado obsoleta. Dada esta situación, las compañías de la industria musical han tenido que adaptarse a las innovaciones de la era digital para mantenerse vigentes. En un principio, las compañías discográficas intentaron detener, sin mucho éxito, las plataformas que distribuían música digital de manera ilegal. Posteriormente, surgieron plataformas legales de música en línea cuyo crecimiento causó el declive de la piratería, salvando de la crisis a la industria de la música. Este crecimiento de plataformas ha generado un aumento gigantesco de la disponibilidad de musica digital. Un ejemplo relevante es Spotify, conteniendo actualmente más de 50 millones de canciones, y crece en 40,000 nuevas diariamente. Otras plataformas como Apple y Amazon Music cuentan con más de 60 millones de canciones.

En estas plataformas se observa una tendencia de los usuarios hacia el descubrimiento de música nueva. Según datos de Spotify, entre 2014 y 2017, el usuario promedio aumentó en un 37 % la cantidad de artistas únicos que escuchaba, pasando de 30 a 41 artistas por semana. Además, se ha observado que los usuarios gustan de descubrir artistas de países diferentes al suyo. Estas observaciones señalan que los usuarios buscan diversificar su escucha. Sin embargo, en catálogos tan inmensos resulta un reto para el usuario descubrir nueva música que realmente sea de su agrado. Por lo tanto, el siguiente paso en la evolución de las plataformas de música en línea podría ser el desarrollo de nuevas maneras de diferenciar sus servicios para ganar usuarios y consolidarse. Una forma de conseguirlo podría ser creando funcionalidades únicas e innovadoras que ayuden a los consumidores a descubrir fácilmente canciones, artistas y listas de reproducción que satisfagan sus necesidades musicales.

Las funcionalidades que actualmente ofrecen las plataformas para ayudar a descubrir nueva música se pueden dividir en dos tipos. El primer tipo crea sugerencias personalizadas por medio de listas de reproducción individuales generadas automáticamente (e.g., descubrimiento semanal en Spotify, o estaciones basadas en una canción o artista de Pandora). El segundo tipo es promover el descubrimiento de música a través de listas de reproducción generales (ej. top diario de éxitos de Spotify) que son generadas mediante el criterio humano y a menudo se utilizan para promocionar canciones y artistas que ya son ampliamente conocidos. Otras listas curadas [1] se enfoca más al descubrimiento de nuevas canciones y artistas. Para que sus colecciones sean accesibles, las plataformas de música deben clasificar la música por género, instrumentación, tempo y estado de ánimo entre otras características. Una taxonomía que comúnmente se utiliza para presentar el contenido es de acuerdo al género musical y/o al estado de ánimo. En algunas ocasiones se incluyen aspectos musicales como tempo e instrumentación. También suelen usarse categorías relacio-

[1] https://es.wikipedia.org/wiki/Curación_de_contenidos

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

nadas con actividades o actitudes (e.g., fiesta, relajación, cena, sueño, viaje, romance). Este tipo de taxonomía trata de esquematizar los aspectos de la vida diaria, sobre todo porque las categorías de música corresponden a las actividades de los clientes objetivo.

A pesar de que ese tipo de clasificación ayuda a organizar la música, algunas veces puede ser ineficaz porque el oyente probablemente quiera descubrir música de acuerdo a una representación de su gusto, emociones y/o personalidad. Algunos autores [1] han sugerido una estructura latente de factores subyacentes a las preferencias musicales independientes del género que refleja principalmente respuestas emocionales/afectivas a la música. También se ha argumentado que existe una relación muy fuerte entre las preferencias por atributos musicales emocionales (e.g., excitación, valencia y profundidad) y la personalidad de las personas [2]. Estas evidencias sugieren el planteamiento de nuevas formas para que las plataformas musicales cambien la taxonomía tradicional que organiza la música según el género y el estilo para explotar estratégicamente el vínculo entre las preferencias musicales, emociones y rasgos de personalidad.

Es por ello que la recuperación de información musical basada en contenido (CB-MIR) ha sido un campo de investigación muy activo en la última década. La CB-MIR tiene muchas aplicaciones como recomendación, etiquetado, indización y catalogación. Las soluciones automáticas al problema del etiquetado automático son invaluables porque hacen que la anotación sea rápida, económica y estandarizada. La emoción es uno de los criterios de búsqueda más importantes para la música [3]. Para desarrollar estas aplicaciones, es esencial que la CB-MIR se enfoque en el etiquetado de canciones basado en patrones emocionales de segmentos vocales y no vocales. Algunas de las funcionalidades que se podrían desarrollar mediante este concepto es navegar utilizando etiquetas afectivas entre colecciones o bases de datos de música y recuperar aquellas canciones que mejor se adapten al estado emocional actual del usuario.

## Taxonomías del estado de ánimo

Para crear un método que obtenga archivos de música de acuerdo a una emoción dada, es necesario adoptar uno o varios modelos emocionales que permitan representar los estado de ánimo de una manera adecuada. Sin embargo, la clasificación de las emociones es muy ambigua debido a muchos aspectos psicológicos relacionados con las emociones de una canción. Existen principalmente dos tipos de modelos para representar emociones, los modelos discretos y los modelos continuos. Los modelos discretos se basan en una lista de adjetivos que representan una emoción (e.g., feliz, triste, enojado). En este tipo de modelo es posible representar el estado de ánimo como un grupo de emociones, en los cuales a su vez, cada uno contiene un subgrupo de emociones. El problema es que esta manera de clasificar emociones es ambigua y dificulta el hacer una estimación concisa del estado de ánimo. Además, en esta clasificación se correlacionan diferentes grupos de adjetivos entre sí. Los modelos de estado de ánimo dimensional, se basan en la afirmación de que los diferentes estados de ánimo están compuestos por combinaciones lineales de un número pequeño de dos o tres estados de ánimo básicos. El modelo más conocido y utilizado es el modelo de afecto circunflejo presentado por Rusell [4]. Este modelo consiste en un espacio bidimensional de placer-disgusto y grado de excitación, lo cual permite identificar etiquetas de emociones como puntos en el espacio del estado de ánimo. Algunos ejemplos de estos puntos en el espacio se pueden ver en la Figura 1. Esta idea la adoptó Thayer [5] y divide ese modelo bidimensional en cuatro cuadrantes, el cual se puede ver como una simplificación del modelo bidimensional de Russell. Ese modelo se muestra en la Figura 2. Estos modelos los diseñan principalmente psicólogos expertos y los han adoptado los investigadores en el área de la CB-MIR. Un enfoque que también se ha probado es el etiquetado social o folcsonomía. En este enfoque varios usuarios colaboran en la descripción emocional de una misma canción asignando palabras claves de un vocabulario y de etiquetas no estructuradas.

**Figura 1.** Modelo de Thayer de cuatro cuadrantes.

## Bases de datos

Para crear bases de datos que puedan ser usadas para el entrenamiento de modelos de clasificación de emociones en música es necesario etiquetar el contenido emocional de un conjunto de canciones. Se obtiene la opinión de varios escuchas de la misma canción obteniendo una etiqueta final mediante votación.

En la Tabla 1 se muestran cinco bases de datos de canciones con etiquetado emocional con sus principales características. Dos de estas bases de datos describen el contenido emocional con base en estados de ánimo. Sin embargo, los estados anímicos algunas veces parecen traslaparse semántica o acústicamente. Las otras dos

bases de datos usan dimensiones emocionales como valencia y excitación para describir el contenido emocional basadas en modelos en modelos psicológicos de las emociones humanas. Este tipo de etiquetado cuando se hace de manera continua facilita la descripción de la evolución emocional de una canción.

Las bases de datos presentadas en la Tabla 1 son las más ampliamente usadas. Algunos foros de evaluación como MediaEval (MediaEval Benchmarking Initiative for Multimedia Evaluation) y MIREX (Music Information Retrieval Evaluation eXchange) las han usado en tareas de evaluación de reconocimiento de emociones en música. Otra base de datos relevante es la descrita en [6] la cual está etiquetadas con excitación y valencia. También existen bases de datos con un enfoque en música de ciertas culturas como canciones Hindi o música de la cultura China.

Con el objetivo de agilizar el procesamiento durante la extracción de características musicales, usualmente no se toma la canción completa sino un segmento representativo de la misma canción (15, 30 o 45 segundos).

## Modelado computacional de emociones

Cuando las etiquetas o clases emocionales correspondientes a las canciones se representan con modelos discretos se dice que es un problema de clasificación. Mientras que la regresión se usa para estimar valores numéricos cuando se etiqueta usando modelos dimensionales. La clasificación de emociones puede convertirse en regresión mediante el uso de un modelo dimensional continuo en lugar de clases. En la literatura es más común encontrar trabajos que usan la clasificación de emociones. También se encuentran trabajos de regresión; sin embargo, estos últimos son menos comunes.

El modelo de Russell es muy utilizado en estudios enfocados en la regresión donde se estima un valor numérico para la valencia y la excitación.

Los trabajos orientados a la clasificación se basan en emociones básicas como feliz, aburrido, enojado y relajado. Otros estudios se basan en los cuatro cuadrantes del modelo de Russell, pero de tal manera que disminuye la complejidad del problema al sólo tener cuatro clases (una clase para cada cuadrante, feliz (Q1), enojado (Q2), triste (Q3) y pacífico (Q4). Mo y Niu [7] realizaron cuatro experimentos: uno con el modelo discreto usando las emociones feliz, triste, temeroso, enojado, sorprendente y tierno, dos con el modelo de Thayer y el último con el modelo de Russell.

Si bien los estudios se basan en el modelo continuo de Russell, se observa que la mayoría de ellos tienen entre 4 y 7 clases emocionales. Es decir, a pesar de que se usa un modelo emocional dimensional, se discretiza la estimación mapeándola a categorías emocionales. Por lo tanto, se podría tratar de extender el número de emociones hacia una representación emocional más flexible.

Los investigadores en esta área han probado varios algoritmos tanto de clasificación como de regresión. El algoritmo más utilizado es SVM (máquina de soporte vectorial), para clasificación y regresión. Además, se han usado redes neuronales feed fordward y convolucionales.

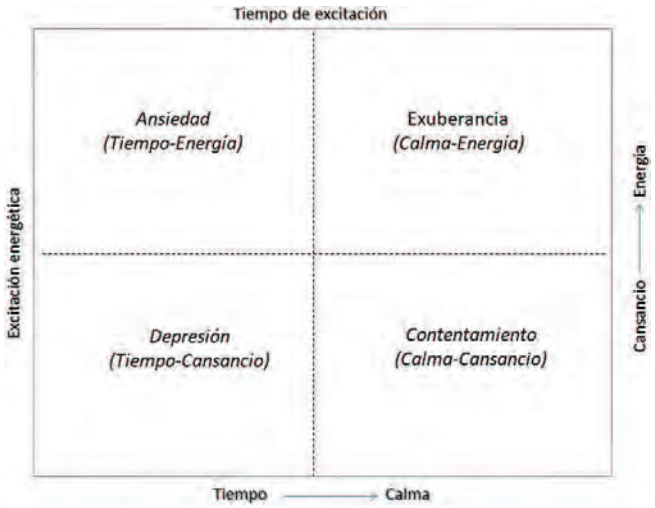



**Figura 2.** Modelo de afecto circunflejo de Russell.

Actualmente, cómo en otras áreas de aprendizaje automático, la tendencia es el uso de redes neuronales profundas. Por ejemplo, en el foro de evaluación Mediaeval "2019 Emotion and Theme Recognition in Music Task", los modelos más usados fueron redes neuronales con diversas arquitecturas que combinan capas convolucionales, GRU (gated recurrent unit), LSTM (long short-term memory), y modelos preentrenados. Las capas convolucionales imitan la habilidad de la corteza visual de nuestro cerebro para identificar objetos sin que afecte su orientación o posición en la imagen. Las capas GRU y LSTM son eficientes para trabajar con datos secuenciales ya que recuerdan las salidas anteriores como entrada. Algunos de los métodos propuestos usaron modelos preentrenados con arquitecturas como MobileNetV2, VGG16, y ResNeXt con bases de datos de imágenes, en particular ImageNet. Dichas arquitecturas son redes neuronales profundas con diferentes combinaciones de tipos de capas y diferentes números de capas. En estos métodos, la representación del audio es por medio de espectrogramas de mel, por lo que los modelos preentrenados son usados para caracterizar los espectrogramas. Después, esta caracterización se pasa a capas posteriores, normalmente densas que se encargan de hacer las estimaciones emocionales. En este mismo foro, también algunos métodos usaron arquitecturas como vgg-ish y VQ-VAE. Estas arquitecturas se preentrenaron con bases de datos de audio como the million song dataset, el MagnaTagATune (MTAT) dataset y el Audioset dataset.

Tabla 1. Bases de datos de 1,000 o más canciones con etiquetado de emociones

| | MTG Jamendo | Mirex 2007 Dataset | Free Music Archive | NTWICM Dataset | AMG1608 Dataset |
|---|---|---|---|---|---|
| Canciones | 18,486 | 1,250 | 1,000 | 2,648 | 1,608 |
| Modelado | Discreto | Discreto | Dimensional | Dimensional | Dimensional |
| Etiquetas | 59 estados de ánimo o tema | 5 grupos de estados de ánimo | Excitación y valencia | Excitación y valencia | Excitación y valencia |
| Anotadores | El autor etiqueta su canción | 2 o 3 por cada canción | Mínimo 10 por canción | 4 por canción | 665 por canción |
| Segmentos | Ninguna | Clips de 30 segundos | Clips de 40 segundos | Ninguna | Clips de 30 segundos |
| Géneros | 19 géneros como: indie pop, rock, clásica, jazz, folk, blues, electrónica | Géneros variados desde música clásica hasta rock y bandas indie actuales | Géneros variados: blues, electrónica, clásica, pop folk, jazz, country, rock | Éxitos comerciales de Reino Unido, pop rock, R&B y dance | Géneros variados:rap, reggae, new age, jazz, pop/rock, latin, folk |

## Descriptores musicales y metadatos

Uno de los retos en la recuperación de archivos de música es la elección de los descriptores musicales adecuados. Algunos descriptores se pueden obtener de los metadatos del archivo digital (e.g., artista, título de la canción, álbum, año, letra). Algunos otros descriptores se pueden calcular a partir del procesamiento de la señal de audio, lo cual se conoce como recuperación basada en el contenido. Una manera de catalogar los descriptores acústicos obtenidos del audio es a partir de las características de alto y bajo nivel. Las de bajo nivel se calculan a partir de la forma de onda del audio como amplitud y espectro. A partir de estas, se pueden derivar características más abstractas conocidas como características de alto nivel. Estas últimas describen conceptos como ritmo, armónicos, timbre, entonación, y tempo. También se identifican algunas características denominadas de nivel medio, como la estructura rítmica y de acordes de una pieza musical.

Para diseñar sistemas de aprendizaje automático que reconozcan emociones es importante seleccionar las características musicales más adecuadas. Un ejemplo de un conjunto de características diseñado para capturar el estado emocional percibido en la música se puede observar en [6]. Este conjunto cubre cinco características de alto nivel: sonoridad, tono, ritmo, timbre y armonía.

- El volumen o sonoridad es la intensidad perceptiva del sonido. Depende principalmente de la intensidad física del sonido, pero también está relacionado con otras propiedades físicas del sonido, como la frecuencia y la duración.
- El tono es el atributo auditivo del sonido que se puede ordenar en una escala de bajo a alto. La música occidental tiene 88 diferentes entonaciones, usualmente denotadas como A0 a C8.
- El ritmo es un rasgo musical importante que representa el pulso de una pieza musical.
- El timbre a menudo se refiere a la calidad o textura del sonido que distingue un sonido musical de otro, incluso cuando los sonidos comparten el mismo tono y volumen.
- La armonía se refiere a la relación entre dos o más tonos simultáneos en la música.

Hoy en día existen diferentes sistemas informáticos que permiten extraer descriptores acústicos de bajo nivel. Por ejemplo, el software OpenSmile esta compuesto por un conjunto de herramientas creado por *Munich Open Speech and Music Interpretation*, el cual permite extraer características acústicas para aplicaciones de procesamiento de señales y aprendizaje automático. Su enfoque es principalmente en las características de bajo nivel del audio y se enfoca al procesamiento en tiempo real. Este software también se puede usar en modo por lotes para el procesamiento de grandes conjuntos de datos. Otra herramienta es MIR Tool Box, el cual es un software ampliamente usado en el área de recuperación de información musical para extraer características musicales como tonalidad, ritmo, estructuras, entre otras. Essentia es otra herramienta para análisis de audio y recuperación de información musical. Además de extraer una gran variedad de descriptores musicales espectrales, temporales, tonales y de alto nivel, incluye herramientas para inferencia con modelos de aprendizaje profundo.

## Recuperación de la información musical basada en emociones

En la actualidad, la investigación en recomendación musical está experimentando un creciente interés por generar recomendaciones que consideren los conceptos psicológicos de personalidad y afecto (estado de ánimo y emoción). Estos conceptos estarán cada vez más integrados en los nuevos prototipos. La motivación para usar estos conceptos es que mientras se escucha música, se ha demostrado que los rasgos de personalidad y los estados afectivos influyen fuertemente en las preferencias musicales. En general, la mayor parte del avance logrado en esta área se concentran en la clasificación y regresión del estado de ánimo en la música. Los siguientes pasos se dirigen hacia la integración de los modelos de clasificación o regresión en sistemas de recomendación y recuperación de información musical y sobre todo en la interfaz de usuario. Un esfuerzo en este sentido se reportó en [8], donde se generó una lista de reproducción basada en el estado de ánimo actual y el comportamiento de la persona, el cual lo captura la aplicación mediante la cámara de un teléfono inteligente. Este sistema trabaja en tiempo real. Sin embargo, el número de canciones incluidas en la base de datos utilizada es muy limitado, sólo 48. En el traba-

jo de Vad et al. [9] se propone una interfaz de usuario que asiste en la creación de listas de reproducción. En la interfaz, el usuario dibuja una trayectoria y selecciona un área en un mapa bidimencional relacionado con descriptores emocionales. El sistema de recomendación de música Moodplay [10] filtra el contenido con base en el estado de ánimo mediante una interfaz interactiva. Este sistema permite al usuario explorar una colección de música mediante dimensiones afectivas. Además, el sistema mezcla la entrada del usuario en el momento de la recomendación con predicciones basadas en un perfil de usuario preexistente.

Las plataformas consolidadas de música en línea ven este tipo de nuevas interfaces como estratégicas para mantenerse vigentes. Un ejemplo relevante es la patente concedida a Spotify en febrero del 2020, en la que se describe un asistente por voz que detecta las emociones en las voces de los usuarios. La patente describe un dispositivo de reproducción capaz de escuchar las reacciones emocionales en el habla del usuario. Este dispositivo deja de reproducir canciones que parezcan evocar enojo, tristeza u otras emociones negativas cuando la voz de un usuario perece triste y recomendarle otro tipo de música. Otra patente concedida a Spotify recientemente describe un sistema que detecta emociones en datos de EEG del usuario. Este sistema responde recomendando una canción particular asociada con la misma emoción que el usuario experimenta.

## Análisis de emociones en música basado en un modelo híbrido

En el Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE) se está trabajando en el desarrollo de un sistema para la recuperación de canciones que analice las propiedades acústicas de la señal de audio. Se pretende que este sistema regrese al usuario las canciones más relevantes que satisfagan sus necesidades de acuerdo a criterios de búsqueda emocionales. Una de las características importantes de este sistema es la utilización de una representación emocional híbrida, es decir, que use una representación dimensional de emociones para caracterizar las propiedades musicales. Este sistema también usa una representación discreta que muestra categorías con las que los usuarios puedan sentirse familiarizados.

Este desarrollo toma como punto de partida la base de datos NTWICM, la cual cuenta con 2,648 canciones que originalmente están etiquetadas con las dimensiones emocionales valencia y excitación. Esta base de datos se ha enriquecido añadiendo las letras de las canciones, y 23 categorías relacionadas con estados de ánimo (e.g., anhelante, cool, optimista). También se le han agregado otros metadatos relevantes de la canción como género, subgénero, ambiente, época, país de origen. Estos datos adicionales fueron obtenidos de la aplicación “MusicID”. Esta aplicación reconoce las canciones escuchadas en el ambiente y genera sus metadatos. Estos datos están disponibles para compartirse y usarse con fines de investigación.

Actualmente se está trabajando en el desarrollo de arquitecturas basadas en redes neuronales profundas tanto para la clasificación de categorías emocionales como para la regresión de niveles de valencia y excitación. Este trabajo ha obtenido resultados alentadores en una red neuronal convolutiva recurrente (CRNN) al usar los datos NTWICM. En la clasificación de las 23 categorías emocionales se obtuvo 0.79 ROC-AUC. En la clasificación de 5 niveles de excitación y valencia se obtuvo 0.78 y 0.83 ROC-AUC, respectivamente.

## Conclusiones

Actualmente existe un gran auge por el desarrollo de nuevas maneras de ayudar a los usuarios de las plataformas de música en línea para descubrir la música que realmente satisfaga sus requerimientos. Una tendencia es tomar en cuenta aspectos afectivos, por lo que las interfaces de recuperación de música deben ofrecer medios flexibles para que los usuarios puedan expresar sus deseos musicales en términos afectivos. La investigación en este sentido puede considerarse aún en una etapa temprana y existen muchos problemas por resolver. A continuación se listan algunos de ellos:

- Muestreo óptimo de la canción. Normalmente, para determinar la carga emocional no se toma en cuenta la duración completa de las canciones si no solamente ciertos segmentos. Se usan extractos breves o los extractos más largos se dividen en segmentos cortos para analizar los resultados. Existe una gran posibilidad de que cambie la emoción percibida por el oyente, ya que la emoción dentro de una canción puede variar con el tiempo.

- Tamaño de las bases de datos. Hasta el momento se ha trabajo con bases de datos relativamente pequeñas. Muchas contribuciones en el área se han hecho con bases de datos de unos pocos cientos de canciones.

- Alternativas al etiquetado emocional. El etiquetado de emociones es subjetivo y costoso ya que es necesario que las mismas canciones estén etiquetadas por varias personas.

- Mejorar la representación emocional. Muchos estudios se basan en el modelo de cuatro cuadrantes de Russell; sin embargo para simplificar la tarea, la clasificación se hace con un número limitado de estados de ánimo.✷

## REFERENCIAS


1. Rentfrow, P., Goldberg, L., y Levitin, D. (2011). The structure of musical preferences: a five-factor model. *Journal of personality and social psychology*, 100(6), 1139.

2. Greenberg, D., Kosinski, M., Stillwell, D., Monteiro, B., Levitin, D., y Rentfrow, P. (2016). The song is you: Preferences for musical attribute dimensions reflect personality. *Social Psychological and Personality Science*, 7(6), 597-605.

3. Aljanaki, A., Yang, Y., y Soleymani, M. (2014). Emotion in Music Task at MediaEval. En *MediaEval*.

4. Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39,(6), 1161-1178.

5. Thayer, R. (1990). The biopsychology of mood and arousal. Oxford University Press.

6. Nawaz, R., y Nisar, H., and Voon, Y., y Yee, T., (2018). Acoustic feature extraction from music songs to predict emotions using neural networks. *Institute of Electrical and Electronics Engineers Inc.*, 166-170.

7. Widowati U., Nugroho S., y Shidik, S. (2018). Classification of Music Moods Based on CNN. *2018 International Seminar on Application for Technology of Information and Communication*, 318-321.

8. Karthik, S. Manasi, A., y Megala, S. (2017). EMOSIC- An Emotion Based Music Player For Android. *2017 IEEE International symposium on signal processing and information technology (ISSPIT)*, 371-376.

9. Vad, B., Boland, D., Williamson, J., Murray-Smith, R., y Steffensen, P. (2015). Design and evaluation of a probabilistic music projection interface. En *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR 2015)*, 134–140.

10. Andjelkovic, I., Parra, D., y O'Donovan, J. (2016). Moodplay: Interactive mood-based music discovery and recommendation. En *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, 275-279.


## SOBRE LOS AUTORES

**Mónica Muñoz-Ramírez** es Ingeniero en Tecnologías de la Información por la Universidad Politécnica de San Luis Potosí. Actualmente es estudiante de tiempo completo de la maestría en ciencias de la computación en el Centro de Investigación Científica y de Educación Superior de Ensenada, en Baja California. Entre sus intereses de investigación destacan la aplicación de algoritmos de aprendizaje automático para la resolución de problemas de diversa índole. Entre sus pasatiempos favoritos se encuentran la música y los idiomas.

**Humberto Pérez-Espinosa** recibió el grado de Licenciado en Ciencias Computacionales en la Benemérita Universidad Autónoma de Puebla (BUAP) en 2004, el grado de Maestro en Ciencias Computacionales en 2006, y el grado de Doctor en Ciencias Computacionales en el Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE) en 2013. Sus intereses de investigación incluyen el aprendizaje automático, computación afectiva, computación paralingüística y el análisis inteligente de audio. Se desempeña como investigador titular en la unidad Nayarit del CICESE (CICESE-UT3) y es miembro del Sistema Nacional de Investigadores nivel uno.

**José Alberto Fernández Zepeda** Obtuvo los grados de Ingeniero Mecánico Electricista y Maestro en Ingeniería Eléctrica en la UNAM, en 1991 y 1994, respectivamente. Obtuvo el grado de Doctor en Filosofía en la Universidad Estatal de Luisiana en 1999. Desde el 2000, se desempeña como investigador titular en el Departamento de Ciencias de la Computación del CICESE. Sus intereses de investigación se centran en el análisis y diseño de algoritmos.

**Juan Martínez-Miranda** es Ing. en Sistemas Computacionales por el Instituto Tecnológico de San Luis Potosí, tiene una maestría en Inteligencia Artificial por la Universidad Politécnica de Cataluña y un doctorado en Ingeniería Informática por la Universidad Complutense de Madrid. Ha trabajado como investigador en el Parque Científico de Barcelona, el Instituto Austriaco de Inteligencia Artificial, la Universidad Politécnica de Valencia y actualmente es investigador titular en la unidad Nayarit del CICESE (CICESE-UT3). Sus intereses de investigación incluyen los sistemas multi-agentes, interacción hombre-máquina y computación afectiva.

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

ARTÍCULO ACEPTADO

# Análisis de técnicas inteligentes para la predicción del flujo de tráfico urbano

Boris Medina Salgado, Eddy Sánchez-DelaCruz, Pilar Pozos-Parra y Javier E. Sierra


## Resumen

En las últimas décadas, el desarrollo de la infraestructura de transporte ha tenido un gran desarrollo, sin embargo, los problemas de tráfico continúan extendiéndose debido al aumento de la población en las zonas urbanas que requieren el uso de estos medios. En América Latina, el crecimiento desordenado de las ciudades aumenta las distancias y recorridos, incrementando la cantidad de autos y motocicletas que tienden a agravar el problema. En este sentido, los sistemas de transporte inteligentes son una alternativa para mejorar el entorno del tráfico. Sin embargo, el procesamiento y modelado de los datos de tráfico en estos sistemas, constituye un desafío debido a la complejidad de la estructura de la red de carreteras, las dependencias espacio-temporales entre ellas y los patrones de tráfico heterogéneos. En este trabajo se presenta una revisión del estado-del-arte de técnicas empleadas para el procesamiento de datos de tráfico urbano y propone una taxonomía general para enfrentar los desafíos aún existentes, donde el conocimiento de flujo de tránsito converge desde un enfoque computacional.


## Introducción

Actualmente, los centros poblados concentran más de la mitad de la población mundial, y se espera que esta proporción siga en incremento. Lo anterior, ha acentuado la problemática de los sistemas de transporte, destacando los problemas relacionados con el control de la congestión que representa una afectación directa en los ciudadanos, tales como: contaminación del aire, consumo de combustible, violación de las reglas de tránsito, contaminación acústica, accidentes y pérdida de tiempo. [1]. A pesar de las numerosas medidas adoptadas para disminuir las graves consecuencias del crecimiento del número de vehículos, el problema sigue existiendo, lo que obstaculiza la velocidad de la urbanización, especialmente en algunas ciudades[1].

La predicción precisa del flujo de tráfico puede aliviar la propagación de la congestión del tráfico, lo que reduce los costos de combustible y la contaminación del aire [2], impactando positivamente en la sostenibilidad de las ciudades. América Latina no es ajena a la situación global, evidenciado en el crecimiento desordenado de las áreas periféricas de las ciudades que aumenta las distancias y los recorridos. De acuerdo a Tomtom Traffic Index, 2019, Bogotá, Lima y la Ciudad de México, se encuentran clasificadas como las ciudades con mayor tráfico vehicular, con niveles de congestión que superan el 50 % de tiempo extra en cada viaje [3].

Los sistemas de transporte están evolucionando para abordarse desde el ámbito de la optimización, de tal forma que puedan enmarcarse dentro del concepto de ciudades inteligentes.

- Los Sistema de Transporte Inteligente (ITS, del acrónimo en Inglés, *Intelligent Transportation Systems*) se perfilan como una alternativa eficiente para mejorar el entorno del tráfico. ITS incorpora tecnologías avanzadas que incluyen tecnologías de sensores electrónicos, tecnologías de transmisión de datos y tecnologías de control inteligente en los sistemas de transporte, para brindar mejores servicios a los conductores y pasajeros en los sistemas de transporte.

- Recientemente, el Internet de las Cosas (IoT, del inglés *Internet of Things*) se ha convertido en un área de investigación emergente dentro de los sistemas que emplean tecnologías de la información y la comunicación. Esta tecnología conecta dispositivos y sensores inteligentes a los que se puede acceder a través de Internet. Los objetos de IoT conectados a Internet envían y reciben datos a través de varios medios de comunicación como Wi-Fi, RFID, WSN, NFC, *Bluetooth*, entre otros. IoT está floreciendo debido a su confiabilidad, interoperabilidad y capacidades de interconexión en tiempo real para aplicaciones de redes inteligentes, además el número de dispositivos conectados sigue en aumento, y con ello, la cantidad de datos recopilados por esos dispositivos también está aumentando. Muchos desarrollos recientes en el análisis de tráfico dependen de las conexiones a dispositivos de IoT.

- Con el rápido desarrollo en el poder computacional, así como el crecimiento en el volumen de datos de tráfico, gran parte de los trabajos se centra en métodos de aprendizaje automático para la previsión de tráfico. A través del aprendizaje automático, estrechamente relacionado con la inteligencia artificial, los sistemas informáticos aprenden a realizar tareas como clasificación, agrupamiento, predicciones, reconocimiento de patrones, entre otras. El aprendizaje automático y profundo es cada vez más popular para la tarea de predicción del flujo de

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

tráfico, debido a su alta capacidad para capturar relaciones complejas no lineales, superar los métodos estadísticos y manejar mayores volúmenes de datos. Actualmente, el surgimiento de los modelos de aprendizaje profundo aprovecha todo el potencial de los enfoques de inteligencia artificial para la predicción de tráfico, con mayor eficacia.

En este documento se realiza una revisión de técnicas de aprendizaje automático para el análisis de datos de movilidad en la predicción del flujo de tráfico en zonas urbanas, filtrando artículos no mayores de cinco años de antigüedad, donde los autores abordaron el uso de algoritmos para la predicción de flujo de tráfico en zonas urbanas, teniendo en cuenta que sus resultados se pudieran cuantificar. Se realiza una categorización de las técnicas usadas y un análisis sintético que destaca ventajas e inconvenientes de éstas, con lo cual proponemos un esquema general de predicción, teniendo en cuenta los desafíos actuales con la finalidad de encontrar resultados de precisión más eficientes.

## Estado-del-arte

La congestión del tráfico se ha convertido en un factor importante que afecta los desplazamientos diarios de las personas y el desarrollo sostenible de las ciudades. Diversos autores han explorado un sinnúmero de técnicas para pronosticar con alta precisión el flujo de tráfico, teniendo en cuenta diferentes aspectos y características relacionadas. De manera general se pueden encontrar tres grandes categorías de modelos de predicción de flujo de tráfico: i) técnicas paramétricas, ii) técnicas de aprendizaje automático y iii) técnicas de aprendizaje profundo. La Figura 1 muestra una taxonomía de las técnicas encontradas en la literatura.

### Técnicas paramétricas

En la última década, se han propuesto diversos métodos para resolver las problemáticas de predicción de flujo de tráfico. (ARIMA, del inglés *Autoregressive Integrated Moving Average*), filtros de Kalman, y modelos de regresión, son las técnicas estadísticas más utilizadas para pronosticar datos de series temporales. Sin embargo, deben entrenarse en un conjunto de datos grande para alcanzar mejores resultados de precisión. Asimismo, los métodos estadísticos tienen muchas suposiciones, escenarios de aplicación ideales y poca precisión de predicción, en comparación con técnicas más recientes. Los métodos clásicos de series de tiempo pueden capturar dependencias temporales en datos de series de tiempo, sin embargo, poseen difcultades para representar la influencia espacial en los problemas de predicción de flujos urbanos [4].

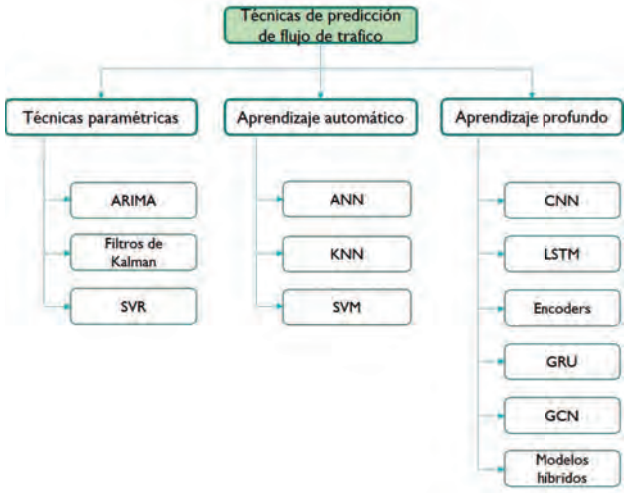



**Figura 1.** Técnicas de predicción de flujo de tráfico urbano.

### Técnicas de aprendizaje automático

Los métodos basados en aprendizaje automático pueden predecir el tráfico futuro con mayor precisión, identificando patrones similares de tráfico basándose en datos históricos, ahorrando tiempo y recursos materiales [4]. Debido a las características estocásticas y no lineales del flujo de tráfico, diversas investigaciones explotan las redes neuronales artificiales para abordar este problema de predicción del flujo de tráfico [2], inicialmente, para extraer patrones de flujo de tráfico (agrupamiento), y luego, basándose en cada patrón, crean un modelo adecuado para predecir el flujo de tráfico.

### Técnicas de aprendizaje profundo

Las técnicas de aprendizaje profundo son una extensión del aprendizaje automático que permiten que los modelos computacionales que se componen de múltiples capas de procesamiento aprendan representaciones de datos con múltiples niveles de abstracción. En las últimas décadas se ha empleado una gran cantidad de sensores para monitorear la situación del tráfico en las vías con la finalidad de controlar las condiciones de tráfico. Esto ha generado una enorme cantidad de datos de alta resolución. Al mismo tiempo, el marco de aprendizaje profundo emergente proporciona otra perspectiva en la extracción de datos con alta dimensionalidad [2]. Estas técnicas han mejorado notablemente los resultados de predicción en el flujo de tráfico vehicular, con altas capacidades para la extracción de dependencias espacio-temporales en conjuntos de datos heterogéneos provenientes de múltiples fuentes. Entre las técnicas de aprendizaje profundo revisadas en el marco de algoritmos de pronóstico de flujo vehicular se destacan: redes neuronales convolucionales (CNN) [5], un tipo de redes neuronales *feedforward* que incluye cálculos de convolu-

ción con una estructura profunda, que pueden capturar con eficiencia la dependencia espacial de la red de tráfico para componer el conjunto de datos de entrada de la serie temporal; las redes neuronales recurrentes cerradas (GRU) [2], empleadas para capturar dependencias temporales porque pueden conservar eficazmente el efecto a largo plazo para los datos de series temporales; las redes de memoria de corto-largo plazo (LSTM) [9], son un modelo mejorado basado en la red neuronal recurrente cuya unidad básica de la capa oculta es un módulo de memoria en lugar de un nodo neuronal tradicional, que incluye una celda de memoria autoconectada y tres unidades de puerta que controlan el flujo de información; *Autoencoders* [8], que constan de un módulo de codificación y un módulo de decodificación diseñados para aprender las propiedades de representación de los datos de entrada; redes convolucionales gráficas (GCN) [6], que también modelan las dependencias espaciales definiendo una estructura de gráfico como la topología física de la red de carretera, y se basa en una extensión de las CNN que operan directamente sobre grafos para aprender las representaciones de capas ocultas de la estructura del gráfico local como las características de los nodos; y los modelos híbridos que combinan dos o más técnicas para aumentar el rendimiento.

## Conjunto de datos

El conjunto de datos más empleado por los autores en sus trabajos es *Caltrans PeMS* [2], [5]. *PeMS* es proporcionado por el Departamento de Transportes de California, EEUU., y recopila constantemente datos de más de 44,681 detectores ubicaciones de autopistas en todo el estado de California, que reportan datos cada 30 segundos. Se encuentran disponibles varios datos de rendimiento, como volúmenes, velocidades, demoras, millas recorridas por vehículos, horas recorridas por vehículos, tiempos de viaje y tráfico diario promedio anual.

Otros conjuntos de datos revisados en la literatura son:

*TaxiBJ* [7], contiene entradas y salidas de taxis en diferentes regiones de Beijing a través de GPS y datos meteorológicos en Beijing, información de identidad del taxi, marca de tiempo, latitud y longitud.

*SZ-Taxi* [6], contiene datos de más de 6 millones de trayectorias de taxi durante 2009, en la ciudad de ShenZhen, China, con atributos como velocidad, ubicación, dirección, estado de ocupancia, tiempo de recorrido.

*Los-Loop* [6], recopila datos mediante detectores de bucle en la ciudad de Los Ángeles relacionados con patrones de volumen de tráfico.

*METR-LA* [5], Conjunto de datos que proporciona el volumen de tráfico diario, a traves de detectores de bucle, en las vías del condado de Los Ángeles, California.

*Taxi-NYC* [9], contiene más de mil millones de registros de taxis de la ciudad de *New York* durante 2015, y proporciona información de ubicación, número de pasajeros, distancia de viaje, tiempos de viaje, entre otros.

*TRAP-2017*, es un conjunto de datos proporcionado por la Policía Nacional de Italia que contiene datos de tráfico, junto con un mapa del segmento de la calle donde se han recopilado los datos, y contiene 111,089,717 de registros que incluyen número de placa, puerta, carril, marca de tiempo y nacionalidad.

## Análisis y nicho de oportunidad

En los últimos años, diversos autores han explorado diferentes técnicas y modelos para la previsión de flujo de tráfico vehicular, empleando desde algoritmos estadísticos como ARIMA, filtros de Kalman y regresiones de soporte vectorial, hasta algoritmos inteligentes donde la red neuronal se emplea de múltiples maneras. Últimamente el desarrollo de los algoritmos de aprendizaje profundo se ha explorado en aplicaciones de movilidad inteligente para la predicción de flujo de tráfico, evidenciado mejoras sustanciales respecto a los algoritmos anteriores. Además, otros autores han explorado modelos híbridos que combinan dos o más técnicas, donde se extraen las ventajas de cada una para potenciar el modelo y aumentar la precisión del pronóstico.

La Figura 2 muestra los resultados más sobresalientes de los algoritmos utilizados para la predicción de flujo de tráfico en el conjunto de datos más usado en la revisión, para tener una mejor aproximación de sus resultados.

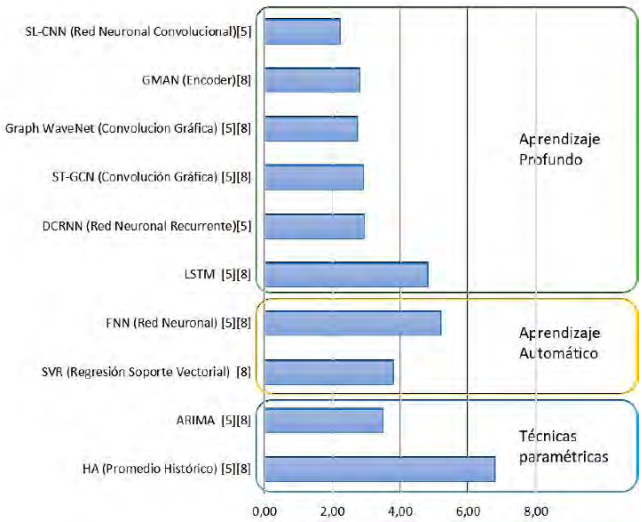



**Figura 2.** Resultados de predicción de tráfico usando métricas MAPE en el conjunto de datos PeMS.

La métrica evaluada MAPE (del inglés, *Mean Absolute Percentage Error*) es un indicador de desempeño que mide el tamaño del error en términos porcentuales. Los métodos basados en datos como ARIMA y el filtrado de Kalman siguen siendo populares, sin embargo, generalmente se basan en la suposición de datos estacionarios,

que a menudo es violada por los datos de tráfico. Debido a las características estocásticas y no lineales del flujo de tráfico, las redes neuronales artificiales (ANN) y el aprendizaje automático explotan mejor las dificultades anteriores.

La Figura 2 muestra que el desempeño de estos modelos estadísticos es inferior a los basados en aprendizaje automático.

En el conjunto de datos *PeMS*, diversos autores coinciden en que el punto clave para el problema de la predicción del flujo de tráfico, es cómo tratar las características de los datos espacio-temporales. Debido a que la secuencia de la característica espacial local histórica describe la información contextual de las series de tiempo, naturalmente afecta la evolución de la siguiente tendencia. Es decir, los puntos de datos de flujo cercanos suelen tener una fuerte relación entre sí.

Además de las características espaciales y temporales obvias del flujo de tráfico, hay varios factores que afectan la variación del volumen de tráfico y la transmisión de la congestión, como las condiciones climáticas, accidentes de tránsito, eventos, puntos de interés, etc. A medida que la fuente de datos se vuelve más heterogénea, el aprendizaje profundo tiende a desempeñar un papel importante, por su capacidad de representar la estructura de percepción multicapa similar a la transmisión de información entre las neuronas del cerebro humano para reconocer patrones de datos no lineales.

Con las ventajas en la extracción de características de imágenes y el aprendizaje recurrente de series de tiempo, los modelos de aprendizaje profundo, como CNN y LSTM, han demostrado un mejor desempeño en la predicción de tráfico con información espacial y temporal. Sin embargo, aunque tanto la dependencia espacial como la dinámica temporal se han considerado en el aprendizaje profundo para la predicción del tráfico, existen desafíos que enfrentar:

- La dependencia espacial puede ser dinámica en el tiempo; la dependencia espacial entre ubicaciones se basa en la similitud del tráfico histórico [10] y el modelo aprende una dependencia espacial estática. Sin embargo, las dependencias entre ubicaciones podrían cambiar con el tiempo.

- Los datos de tráfico no son estrictamente periódicos [10], y pueden variar de acuerdo a otros factores como las condiciones climáticas, accidentes de tráfico o eventos programados.

- Poseen dificultades para funcionar en el contexto de una red de carreteras urbanas con una estructura topológica compleja, por lo que, en esencia, presentan dificultades para describir la dependencia espacial.

Las correlaciones de las condiciones del tráfico entre los sensores en una red de carreteras cambian significativamente con el tiempo. Para abordar las complicadas dependencias espaciales de las redes de carreteras, dado que naturalmente está estructurada como un gráfico en un espacio no euclidiano con carreteras como nodos y sus conexiones naturales como bordes, las investigaciones recientes están empleando la red de tráfico como un gráfico. Las GCN, que han logrado resultados notables en el modelamiento de flujos de tráfico como gráficos espacio-temporales y manejo de correlaciones no euclidianas, se integran con las RNN o las CNN para incorporar el conocimiento previo de la red de carretera y capturar las correlaciones entre nodos emparejados. En la literatura, los esfuerzos recientes han aplicado varios métodos de aprendizaje profundo para el pronóstico del tráfico, por ejemplo, aprovechando GCN para el modelado de dependencia espacial y utilizando RNN para capturar la dinámica temporal [9].

Aunque las redes neuronales de gráficos espacio-temporales han logrado mejoras significativas en comparación con los métodos sin utilizar una estructura de gráfico, como se muestra en la Figura 2, aún existen algunos desafíos por enfrentar:

- Las redes neuronales de gráficos espacio-temporales definen heurísticamente la estructura del gráfico como la topología física de la red de carreteras, ignorando la posible dependencia de la estructura del gráfico sobre los datos de tráfico [6]; por lo tanto, no se garantiza que sea una descripción óptima de la dependencia entre los flujos de tráfico.

- La introducción de incertidumbre en la estructura del gráfico, como la caída aleatoria de nodos o bordes, podría mejorar la capacidad de generalización de los GCN; la estructura del gráfico empleada en los métodos existentes basados en GCN es determinista, lo que significa una falta de investigación de la incertidumbre [6].

- Algunos estudios basados en GCN, además, se centran principalmente en modelar la dependencia espacial utilizando GCN en un gráfico ponderado fijo, sin embargo, las correlaciones entre nodos de carretera son dinámicas y cambian con el tiempo [8].

## Estructura propuesta para la predicción de flujo de tráfico

La Figura 3 muestra un esquema general para la predicción del flujo de tráfico vehicular, que consta esencialmente de tres etapas: recolección de datos, preprocesamiento y una estructura de red neuronal para el procesamiento y análisis predictivo del flujo de tráfico, princi-

palmente profundo, dados los notables resultados frente a otras técnicas como estadística o de aprendizaje automático.

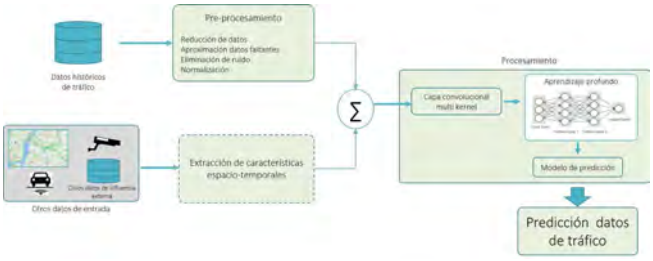



**Figura 3.** Esquema propuesto para la predicción de flujo de tráfico.

En la recolección de datos encontramos los datos de entrada que se consideran para pronosticar la condición del tráfico, dentro de los cuales se encuentra información espacio-temporal sobre el flujo de tráfico actual e histórico del área urbana de estudio, proporcionada por redes de sensores, detectores en la vía, información enviada por GPS de los vehículos, entre otros. Las fuentes de datos de influencia externa, como información meteorológica, condiciones de la carretera, programación de eventos en zonas de influencia, puntos de interés, información de accidentes y otros, pueden ayudar a los resultados de predicción del algoritmo.

Los datos recopilados y transformados se procesan previamente para refinar puntos de datos no estructurados, desequilibrados o no estandarizados; el ruido se elimina para ayudar a reducir el tráfico y la latencia de la red, y se normaliza para escalar los datos de acuerdo con la distribución normal estándar.

El esquema de procesamiento analiza la información utilizando algoritmos de análisis de datos impulsados por modelos de aprendizaje automático para extraer la información clave requerida. Los algoritmos de aprendizaje profundo pueden extraer dependencias espaciales y temporales de manera más eficiente. Una función de núcleo múltiple puede asignar diferentes pesos a las unidades neuronales y podría ayudar a modelar el rango de dependencia espacial entre regiones vecinas que pueden ser diferentes. Asimismo, un algoritmo que permita inferir la incertidumbre incorporada por los datos de flujo de tráfico externos podría incrementar el rendimiento del modelo.

Finalmente, a través de un servicio de integración de datos unificado, se generan enlaces de información, de tal manera que los conjuntos de datos fusionados, de múltiples fuentes, están preparados para ser consultados en búsqueda de información.

## Conclusiones

En este documento realizamos una revisión de técnicas inteligentes para el análisis predictivo del flujo de tráfico en áreas urbanas. Se identificaron y categorizaron técnicas recientemente utilizadas en la literatura. Describimos brevemente, mediante un análisis sintético, los alcances y limitaciones de los modelos que actualmente se emplean en el pronóstico de flujo de tráfico vehicular. A partir de ahí podemos decir que tanto modelos basados en aprendizaje profundo como algoritmos híbridos logran resultados prometedores para la predicción del tráfico urbano, superiores a modelos tradicionales como ARIMA, SVR y algoritmos inteligentes KNN y SVM. Entre ellos, GCN ha logrado resultados notables por su capacidad para abordar las complicadas dependencias espaciales de las redes de carreteras, dada su estructura gráfica natural. Los esfuerzos recientes han aprovechado GCN para extraer características espaciales y RNN para extraer dinámicas temporales, sin embargo, aún existen desafíos que podrían abordarse para mejorar los resultados de la predicción, teniendo en cuenta que se pueden abordar aspectos en la captura y extracción de dependencias. Datos externos espacio-temporales que afectan las condiciones de movilidad, como clima, eventos, puntos de interés y redes sociales. Asimismo, considerar que el rango de dependencia espacial puede ser diferente entre regiones espacialmente próximas o vecinas, cuyas características varían con la condición del tráfico, y puede ayudar a proporcionar información dinámica sobre la red de tráfico urbano.✳

## REFERENCIAS


1. Xu, T., Han, G., Qi, X., Du, J., Lin, C., y Shu, L. (2020). A hybrid machine learning model for demand prediction of edge-computing-based bike-sharing system using internet of things. *IEEE Internet of Things Journal*, 7(8), 7345-7356.

2. Zhang, D., y Kabuka, M. (2018). Combining weather condition data to predict traffic flow: a gru-based deep learning approach. *IET Intelligent Transport Systems*, 12(7), 578–585.

3. TomTom (2016). TomTom traffic index. Recuperado el 21 de agosto de 2019, de `https://www.tomtom.com/en_gb/traffic-index/`.

4. Xie, P., Li, T., Liu, J., Du, S., Yang, X., y Zhang, J. (2020). Urban flow prediction from spatiotemporal data using machine learning: A survey. *Information Fusion*, 59, 1–12.

5. Zhang, Q., Chang, J., Meng, G., Xiang, S., y Pan, C. (2020). Spatio-Temporal Graph Structure Learning for Traffic Forecasting. Wb *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(1), 1177-1185.

6. Fu, J., Zhou, W., y Chen, Z. (2020). Bayesian spatio-temporal graph convolutional network for traffic forecasting. arXiv:2010.07498.

7. Jiang, W. (2021). Taxibj21: An open crowd flow dataset based on beijing taxi gps trajectories. *Internet Technology Letters*, 297.

8. Zheng, C., Fan, X., Wang, C., y Qi, J. (2020). Gman: A graph multi-attention network for traffic prediction. En *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(1), 1234-1241.

9. Liao, S., Zhou, L., Di, X., Yuan, B., y Xiong,J. (2018). Large-scale short-term urban taxi de-mand forecasting using deep learning, En *2018 23rd Asia and South Pacific Design Automation Conference (ASP-DAC)*, IEEE, 428–433.

10. Yao, H., Tang, X., Wei, H., Zheng, G., y Li, Z. (2019). Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. En *Proceedings of the AAAI Conference on Artificial Intelligence*, 33, 5668–5675.

## SOBRE LOS AUTORES

**Boris Medina Salgado** es estudiante del Doctorado en Ciencias de la Ingeniería en el Departamento de Posgrado e Investigación del Instituto Tecnológico Superior de Misantla, Veracruz, México. Boris obtuvo el grado de Maestro en Automatización y Control Industrial en el Instituto Tecnológico Metropolitano de Medellín, Colombia, en el año 2014. Sus intereses de investigación incluyen Análisis de datos mediante algoritmos inteligentes.

**Eddy Sánchez-DelaCruz** es Doctor en Sistemas Computacionales por la Universidad del Sur, Maestro en Sistemas Computacionales y Licenciado en Informática Administrativa por la Universidad Autónoma de Tabasco. Actualmente es Profesor-Investigador en el Departamento de Posgrado e Investigación del Instituto Tecnológico Superior de Misantla, Veracruz, México. El Dr. Sánchez-DelaCruz es miembro del Sistema Nacional de Investigadores, nivel I, miembro de la RedICA-CONACyT, Perfil deseable PRODEP y fundador y líder del Cuerpo Académico Inteligencia Artificial. Sus intereses de investigación son el aprendizaje automático y la ciencia de datos con un enfoque multidisciplinario (`www.eddsac.com`).

**Pilar Pozos-Parra** es Doctora en Inteligencia Artificial (IA) por ISAE-SUPAERO. Es miembro del Sistema Nacional de Investigadores, nivel II. Sus intereses de investigación son la IA explicable y la aplicación de la IA al área de la salud.

**Javier E. Sierra** es Doctor en Ingeniería por la Universidad Pontificia Bolivarian, Medellin, Colombia. Investigador Senior, reconocido por Colciencias, en Colombia. Sus intereses de investigación se destacan en aplicaciones desarrolladas en el marco de las ciudades inteligentes.

Imagen tomada de: `https://www.tomtom.com/blog/traffic-and-travel-information/road-traffic-prediction/`.

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

ARTÍCULO ACEPTADO

# Asistentes inteligentes para el manejo de datos geoespaciales ante desastres naturales

Daniela Moctezuma, Julio Godoy y Diego Seco

En los últimos años, nuestra sociedad se ha enfrentado a un número cada vez mayor de desastres naturales. Inundaciones, olas de calor extremo, incendios, terremotos y huracanes, son sólo algunos de los tipos de desastres naturales que han golpeado diferentes áreas geográficas de nuestro planeta durante la última década, poniendo en peligro la vida de millones de personas. La región latinoamericana, no es una excepción, ya que se le puede considerar un hiper laboratorio de desastres naturales. La recopilación y el intercambio de información geográfica de estos acontecimientos, de manera precisa y oportuna, pueden salvar millones de vidas. Por ejemplo, la información geográfica podría ayudar a las personas que se encuentran en la trayectoria de un huracán a encontrar un refugio cercano y permanecer seguras, lo que podría reducir el impacto general del desastre natural [11].

En particular, gracias a que la Inteligencia Artificial (IA) ha aumentado su popularidad en los últimos años, los gobiernos y la industria han comenzado a utilizar técnicas de IA para procesar la enorme cantidad de datos de que disponen y mejorar su proceso de toma de decisiones. La aparición de Asistentes Inteligentes (AI) ha ayudado a distribuir eficientemente el tiempo en nuestra vida cotidiana; esta nueva forma de interactuar con los dispositivos y los datos está definiendo una nueva era de interacción entre el hombre y la máquina, para la que la inteligencia artificial ha mostrado un enorme potencial. En este sentido, un asistente inteligente puede ser muy útil en muchos dominios, como el entretenimiento, la seguridad, el marketing, etc., y si se combina con el uso de los datos geoespaciales se vuelve un tema de investigación muy interesante, especialmente en situaciones de desastre.

Con relación a este tema, recientemente se creó la red IDEAIS (Asistentes Inteligentes para las Infraestructuras de Datos Espaciales) [12], la cual constituye una red de investigación compuesta por múltiples socios iberoamericanos cuyo objetivo es investigar sobre sistemas inteligentes, especialmente asistentes inteligentes aplicados a la recuperación, manejo y acceso a información geográfica. En esta red participan investigadores de diversas áreas del conocimiento como inteligencia artificial, infraestructura de datos geoespaciales, recuperación de información, interacción humano-computadora, entre otras. Uno de los principales objetivos de esta red es unir esfuerzos para facilitar el acceso a la información geográfica volviendo su acceso más amigable, útil y eficiente, sobretodo cuando un desastre ocurre. El presente artículo, tiene como objetivo mostrar, de forma general, los avances en la literatura respecto a asistentes inteligentes aplicados al uso de información geográfica.

## Principales retos en el desarrollo de asistentes inteligentes aplicados a desastres naturales

**Calidad, cantidad y disponibilidad de los datos**. En los desastres naturales, en muchos casos, es importante obtener datos de varias organizaciones y los datos suelen ser muy heterogéneos; sin embargo, es importante hacer coincidir la información diversa de todas las instituciones para obtener un dato de calidad y utilidad. La disponibilidad de suficientes datos geo-espaciales de buena calidad es el insumo clave de cualquier sistema de gestión de desastres. La disponibilidad de dicha información puede diferir entre los países y entre las zonas de un mismo país.

**Desarrollo de sistemas inteligentes**. El uso de métodos de procesamiento del lenguaje natural (PLN) para el desarrollo de algoritmos que sirvan como medio de comunicación entre el sistema y el usuario, ya sea para obtener información y/o interactuar con la Infraestructura de Datos Espaciales (IDE), es un tema de investigación novedoso, actual y pertinente. Concretamente, los dos principales desafíos en la generación de asistentes virtuales inteligentes para interactuar con la IDE son: i) el diseño de algoritmos inteligentes que utilicen el lenguaje natural con la voz o texto; y ii) el desarrollo de algoritmos capaces de razonar y actuar con la menor interacción humana posible, lo cual es una tarea difícil de resolver [14].

**Accesibilidad**. Los sistemas inteligentes para la gestión de desastres deben proporcionar una interfaz de usuario robusta, que permita acceder a ella desde una variedad de dispositivos y en entornos donde las comunicaciones pueden estar restringidas.

## ¿Qué se ha desarrollado hasta hoy?

En nuestra sociedad, muchas personas recurren a un asistente personal para realizar sus actividades rutinarias, por ejemplo, responder correos, manejar calendarios, etc. Hoy, el concepto de asistente virtual inteligente es muy popular y aceptado. Un asistente virtual puede realizar varias tareas como: buscar información en internet, re-

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

producir música, procesar información, etc. En este sentido, Meissner et al. [6] proponen un sistema integrado de comunicación e información en la gestión de desastres, con un caso de estudio en Alemania. Identifican diferentes desafíos en el uso de las tecnologías de la información (TIC) para la respuesta y la recuperación en casos de desastre, como la integración adecuada de la información, el acceso rápido a los datos, la actualización oportuna de la información y la normalización de la misma. Los autores proponen una arquitectura para un sistema integrado que aborda varias problemáticas, como el flujo de información entre la oficina central de operaciones y las personas que trabajan en terreno, la recopilación y gestión de datos a diferentes niveles, el calendario de recursos, entre otras. En general, los autores llegan a la conclusión de que, para aplicar el sistema propuesto, varias disciplinas de las tecnologías de la información deben trabajar juntas para proporcionar una solución integrada. Goswami et al. [7] revisan las técnicas de minería de datos utilizadas en la predicción de los desastres naturales, junto con una descripción del tipo y las fuentes de datos disponibles para los diferentes tipos de desastres. Como estudio de caso, se centran en la India, cuyo territorio está sujeto a un número significativo (y creciente) de desastres naturales por año. Los autores llegan a la conclusión de que, además de las fuentes de datos tradicionales que pueden utilizarse para predecir posibles nuevos desastres, las redes sociales pueden ser una fuente importante para la alerta temprana de este tipo de acontecimientos.

Otra propuesta es SIRENE [8], que es una infraestructura de datos espaciales (IDE) que proporciona acceso a datos geoespaciales distribuidos, de diferentes fuentes, para ayudar a los responsables de la toma de decisiones en la mitigación de los impactos de los desastres naturales (ver Figura 1). Entre otras características, SIRENE recupera información clave de una serie de repositorios para ayudar en las diferentes fases de emergencia de forma colaborativa y modular. Además, ayuda a los principales actores con sugerencias sobre posibles medidas de mitigación.

Se pueden identificar cuatro formas principales de cómo la gente utiliza las redes sociales durante un desastre natural: para conectarse con la familia y los amigos, para obtener información adicional sobre la situación local, para complementar los medios de comunicación tradicionales en su conocimiento del desastre, y para proveerse mutuamente medios para contactar los servicios necesarios en una crisis [9]. Además, se pueden llevar a cabo diversas tareas mediante las redes sociales como ayudar a preparar a los ciudadanos en las zonas que pueden verse afectadas por un desastre, difundir información sobre la situación en tiempo real, movilizar y coordinar los esfuerzos de ayuda temprana, y optimizar las tareas de recuperación.

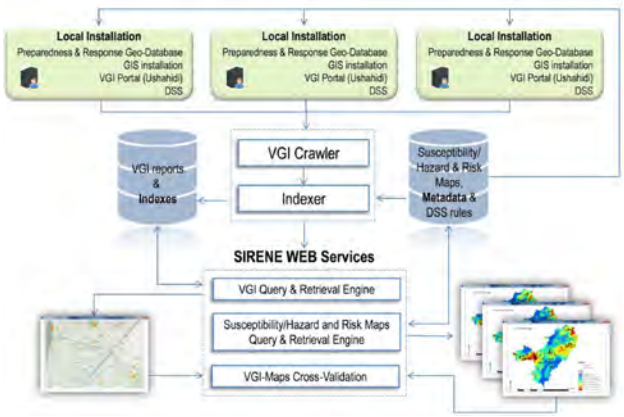



**Figura 1.** Esquema conceptual de los módulos distribuidos que componen SIRENE. Adaptación al español de la figura original publicada en [8].

## Inteligencia artificial para asistentes virtuales

Aunque los asistentes virtuales son una herramienta cada vez más utilizada en diversas áreas, desafortunadamente, estos no son muy populares para el manejo de datos en el dominio de desastres. Por mencionar algunos ejemplos de asistentes virtuales aplicados a diversas áreas, tenemos el caso del asistente virtual utilizado para instrumentos de laboratorio propuesto por Austerjost et al. [2]. En dicho trabajo, el objetivo es establecer un laboratorio con una interfaz basada en voz para controlar los instrumentos del laboratorio. Para lograr esto, se utilizó tecnología de Internet de las cosas (IoT, por su sigla en inglés) para conectar los dispositivos e instrumentos del laboratorio.

Una de las limitaciones de los asistentes virtuales tradicionales es el número fijo (pre-programado) de posibles interacciones. Esto perjudica su uso en ambientes altamente dinámicos e impredecibles, donde el usuario puede querer utilizar el asistente virtual con elementos o situaciones no considerados en su programación. Para abordar este problema, Azaria et al. [10] proponen el uso de herramientas de procesamiento de lenguaje natural para permitir que el asistente virtual aprenda nuevas instrucciones, basado en su explicación mediante primitivas de lenguaje natural. En su trabajo, los autores también muestran cómo el asistente puede generalizar a situaciones similares luego de aprender nuevos comandos.

Sulochana et al. [11] presentan un asistente virtual para el control de electrodomésticos. El asistente es capaz de procesar comandos de voz para un conjunto de tareas como búsquedas en internet, reporte del clima, cálculos básicos, revisar correos electrónicos, etc. Esta propuesta de asistente utiliza técnicas de aprendizaje automático y de procesamiento de lenguaje natural.

Otro asistente virtual inteligente capaz de aprender nuevas tareas mediante lenguaje natural (por medio de

la voz) es propuesto por Merav et al. [12]. Este método utiliza un agente de aprendizaje por instrucción (llamado *LIA*), compuesto por varios elementos inteligentes que permiten resumir textos de forma automatizada, detección y reconocimiento automático del habla, generación de lenguaje natural, etc. Los autores utilizan Google ASR (Automatic Speech Recognition), un parser CCG (Combinatory Categorical Grammar)[1], y un conjunto de reglas lógicas para aprender nuevos comandos.

Otro asistente virtual (llamado *SEVA*) se propuso por Krishnan et al. [13]. Este asistente ayuda al ingeniero de sistemas en su trabajo diario, a través de un método de búsqueda de respuestas usando procesamiento del lenguaje natural, basado en entradas de texto. El sistema *SEVA* tiene un módulo de procesamiento de lenguaje natural que utiliza un diccionario y sistemas de normalización de verbos, extracción de reglas y lenguaje de consultas. El motor de inferencia trabaja con un módulo de razonamiento para lenguaje de ontología, además de un módulo de razonamiento temporal y de reglas. Como salida, *SEVA* tiene tres opciones: una respuesta, una pregunta (en modo conversacional) y respuestas intermedias del sistema.

Como se puede observar, hay diversas aplicaciones para asistentes virtuales inteligentes, la mayoría de las cuales están relacionadas con actividades del hogar, educación, manipulación de instrumentos, entretenimiento, etc. Sin embargo, es importante señalar que, al menos en la revisión que llevamos a cabo, existen pocos trabajos sobre asistentes inteligentes aplicados al manejo de desastres naturales.

## Asistentes virtuales inteligentes para la gestión de desastres

Si hablamos específicamente de asistentes virtuales inteligentes para la gestión de desastres, son pocos los trabajos publicados que abordan el problema completo, con cada uno de sus componentes y aplicaciones. Tal es el caso del trabajo presentado por Tsai et al. [14], donde analizan el desempeño de sistemas de chatbots con datos relacionados a desastres. Este sistema es llamado *Ask Diana* y es capaz de interactuar con el usuario mediante lenguaje natural por medio usualmente de texto. *Ask Diana* tiene tres objetivos principales. Primero, el sistema debe considerar una base de datos de gran tamaño para cualquier proceso de toma de decisiones en desastres relacionados con el agua. Segundo, el sistema debe analizar automáticamente los datos o información que el usuario solicite. Y tercero, el sistema debe proveer una interfaz simple e intuitiva para operar y llevar a cabo todos los procedimientos incluidos. Los autores aplicaron un clasificador basado en árbol de decisión para la gestión y clasificación de los datos de desastres relacionados con el agua. Este clasificador representa un atributo de los datos en cada nodo; cada nodo hoja del árbol representa un único dato y, finalmente, el camino completo desde la raíz hasta la hoja representa la regla de clasificación.

Otro asistente inteligente que opera con base en a preguntas sobre inundaciones fue propuesto por Sermet y Demir [15]. Este sistema inteligente utiliza reconocimiento de voz por medio de herramientas de procesamiento de lenguaje natural, y se basa en una ontología generalizada. El sistema completo, llamado *Flood AI*, incluye un sistema web, chatbots, aplicaciones para teléfonos inteligentes y dispositivos para hogares inteligentes. En este sistema, el motor de conocimiento usa las ontologías y los conceptos para conectar el comando del usuario con la información relevante. La comunicación es a través de datos de voz y texto para permitir interacción y presentar la información solicitada a los usuarios. Este sistema fue desarrollado con base a un motor de generación de conocimiento que acepta consultas en lenguaje natural para analizar grandes conjuntos de datos relacionados con inundaciones y, a continuación, entrega la información resultante al usuario. En particular, para el módulo de procesamiento de lenguaje natural varias herramientas fueron evaluadas (por ejemplo, Stanford NLP, NlPTools); sin embargo, los autores indican que estas herramientas consumen demasiado tiempo para un sistema basado en web, por esta razón, desarrollaron una herramienta de procesamiento de lenguaje natural basado en ontologías. Para el proceso de validación, un grupo de personas interesadas, relacionadas con ciencias ambientales, interactuaron con el sistema y entregaron su retroalimentación. En general, se concluye que el sistema es completo y tiene un buen desempeño en la gestión de información de inundaciones.

Los trabajos anteriormente descritos, son ejemplos recientes donde los asistentes han sido desarrollados para ayudar a los humanos a coordinar los recursos de manera explícita o implícita, luego de la ocurrencia de un desastre.

## Partes o elementos de un asistente inteligente para manejo de datos geoespaciales

Si bien, ya hemos descrito algunos trabajos que abordan, parcial o totalmente, el concepto de asistente inteligente para el manejo de datos geoespaciales, aplicados directamente ante desastres naturales, ahora veremos qué partes son esenciales para ello, desde un punto de vista conceptual enmarcado dentro del proyecto IDEAIS [3], anteriormente mencionado.

La Figura 2 muestra este modelo conceptual que utiliza varios tipos de datos como texto, voz, etc. y, a través de algoritmos de inteligencia artificial, este asistente virtual es consultado por el responsable de tomar decisio-

[1] http://groups.inf.ed.ac.uk/ccg/software.html

nes. Luego de las tareas de procesamiento de los datos, el asistente entrega al usuario la respuesta en forma de gráficos, mapas, datos, etc.

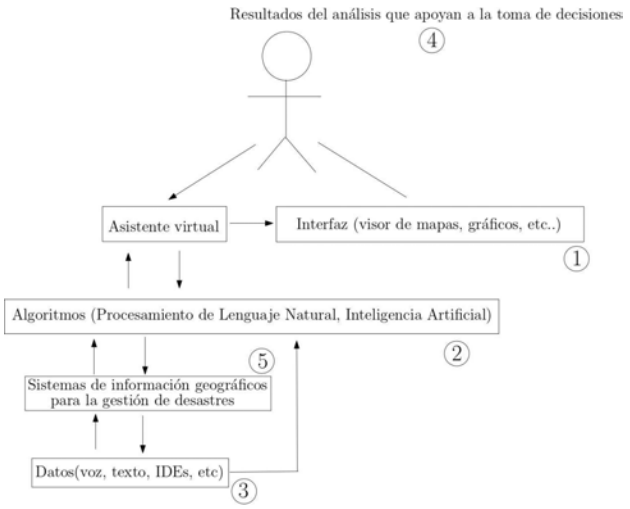



**Figura 2.** Esquema general y conceptual del asistente inteligente propuesto en IDEAIS.

Para este concepto de asistente virtual inteligente, es necesario tomar en cuenta los trabajos revisados en las secciones previas. La Figura 2 muestra cada componente esencial de este modelo conceptual, enumerado del 1 al 5. Es importante mencionar que este concepto surge dentro de la red IDEAIS de la cual forman parte los autores del presente trabajo, para mayor información consultar el trabajo de Bernabé et al. [3]. Este diseño conceptual consiste en el uso de datos geoespaciales desde varias fuentes y mediante sistemas de información, para permitir su análisis en apoyo al proceso de toma de decisiones. Este análisis puede realizarse mediante algoritmos de inteligencia artificial, por ejemplo de las áreas de procesamiento de lenguaje natural, aprendizaje automático, etc., usando voz, texto o una combinación de ambos.

Finalmente, con el propósito de ubicar a cada uno de los trabajos revisados y descritos anteriormente, dentro de las áreas o partes del modelo conceptual reflejado en la Figura 2, en la siguiente lista indicamos cuáles pueden ser de ayuda en cada uno de estos componentes:

1. Procesamiento de lenguaje natural - aprendizaje automático. Amos et al. en [10], Austerjost et al. en [2], Sulochana et al. en [11], Merav et al. en [12], Krishnan J. et al. en [13].

2. Uso de datos, voz, texto. Azaria et al., en [10], Austerjost et al. en [2], Sulochana et al. en [11], Merav et al. en [12], Krishnan J. et al. en [13].

3. Análisis de desastres naturales para toma de decisiones. Tsai M. et al. en [14], Sermet Y. et al. in [15].

4. Sistemas de información para gestión de desastres. Velev D. et al. en [9], Meissner A. et al. en [6], Sterlacchini S. et al. en [8], Goswami S. et al. en [7]

Como se puede observar en esta lista, hay diversos trabajos relacionados al uso del procesamiento de lenguaje natural y métodos de inteligencia artificial en general. Si bien los avances son significativos, aun falta conjugar los diversos esfuerzos hechos en diferentes áreas para obtener un sistema inteligente para datos geoespaciales aplicado a desastres naturales.

✷

## REFERENCIAS


1. Meissner, A., Luckenbach, T., et al. (2002). Design challenges for an integrated disaster management communication and information system. *IEEE Workshop on Disaster Recovery Networks (DIREN 2002)*, 1-7.

2. Goswami, S., Chakraborty, S., et al. (2018). A review on application of data mining techniques to combat natural disasters. *Ain Shams Engineering Journal*, 365-378.

3. Sterlacchini, S., Bordogna, G., et al. (2018). SIRENE: A Spatial Data Infrastructure to Enhance Communities Resilience to Disaster-Related Emergency. *International Journal of Disaster Risk Science*, 129-142.

4. Velev, D., y Zlateva, P. (2012). Use of social media in natural disaster management. *Intl. Proc. of Economic Development and Research*, 41-45.

5. Azaria, A., Srivastava, S. et al. (2020). An agent for learning new natural language commands. *Autonomous Agents and Multi-Agent Systems*, 6.

6. Sulochana, D., Zishan, M. et al. (2019). Artificial Intelligence based Personal Assistant. *Asian Journal For Convergence In Technology (AJCT)*.

7. Merav, C., y Amos, A. (2019). LIA: A Virtual Assistant that Can Be Taught New Commands by Speech. *International Journal of Human–Computer Interaction*, 1596-1607.

8. Krishnan, J., Coronado, P., y Reed, T. (2019). SEVA: A Systems Engineer's Virtual Assistant. *AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering*.

9. Tsai, M., Chen, J., et al. (2019). Ask Diana: A Keyword-Based Chatbot System for Water-Related Disaster Management. *Water Journal*, 234.

10. Sermet, Y. y Demir, I., (2018). An intelligent system on knowledge generation and communication about flooding. *Environmental Modelling and Software*, 51-60.

11. Martínez-Álvares, F., y Morales-Esteban, A. (2019). Big data and natural disasters: New approaches for spatial and temporal massive data analysis. *Computers & Geosciences*, 38-39.

12. Bernabé, M.A., Estima, J., et al. (2019). IDEAIS: Smart Voice Assistants to Improve Interaction with SDIs. *ArXiv*.

13. Cutter, L. (2003). GI science, disasters, and emergency management. *Transactions in GIS*, 439-446.

## SOBRE LOS AUTORES

**Daniela Moctezuma** obtuvo su doctorado en Tecnologías de la Información y Sistemas Informáticos por la Universidad Rey Juan Carlos en Madrid, España. Recibió el grado en junio de 2013 con mención honorífica Cum laude. Actualmente es investigadora y profesora (Cátedra Conacyt) del Centro de Investigación en Ciencias de Información Geoespacial (CentroGEO) desde el 2014. Sus principales líneas de investigación son visión artificial, aprendizaje automático, categorización de texto, percepción remota y vídeo vigilancia inteligente.

**Julio E. Godoy** Obtuvo su doctorado en Computación en la Universidad de Minnesota, USA. Actualmente es profesor asistente en la Universidad de Concepción, Chile. Sus intereses de investigación están en la intersección de inteligencia artificial distribuida y robótica.

**Diego Seco** obtuvo el título de Ingeniero en Informática por la Universidad de A Coruña en el año 2006 y el título de Doctor en Computación por la misma universidad en el año 2009. Sus principales intereses de investigación incluyen la recuperación de información geográfica, los sistemas de información geográfica, y las estructuras de datos y algoritmos para la representación y recuperación de información textual y geográfica. Actualmente es profesor asociado en el Departamento de Ingeniería Informática y Ciencias de la Computación de la Universidad de Concepción (Chile).

---

## *Asistentes Inteligentes para las Infraestructuras de Datos Espaciales (IDEAIS)*

El objetivo general de este Proyecto es la creación de un entorno iberoamericano de colaboración que desarrolle un sistema de asistentes inteligentes para facilitar el acceso a la Información Geográfica de las IDE.

Imagen tomada de: `https://imfd.cl`.

ARTÍCULO ACEPTADO

# Implementación de la metodología CRISP-DM para el pronóstico de los precios de la uva sonorense en el mercado de Estados Unidos

Vicente Solis Sandoval, Juan Martin Preciado Rodriguez y Luis Felipe Romero Dessens

**Una de las metodologías más usuales para el desarrollo de este tipo de proyectos es CRISP-DM.**

## Introducción

El desarrollo de las TICs hace asequible la generación y almacenamiento de grandes bases de datos. Sin embargo, una proporción considerable de datos queda en espera ser transformados en conocimiento que genere valor para los interesados [1]. Esto provocado por una parte por la dificultad para el procesamiento de datos dado su volumen y complejidad, así como, la escasez de herramientas y capacidades para su gestión eficiente [1][2]. Con el propósito de reducir la asimetría entre la generación de datos y conocimiento, es que surge el proceso de "Descubrimiento de Conocimiento en Bases de Datos" (KDD) en 1996, definido como el proceso no trivial para identificar patrones en los datos que sean válidos, novedosos, potencialmente útiles y comprensibles, a través de algoritmos de aprendizaje automático incluidos en la Minería de Datos (MD) [2].

El buen logro del objetivo del proceso KDD, va más allá de los métodos y técnicas analíticas utilizadas [3], en gran medida depende de las personas involucradas en el proyecto. Lo que lleva a la necesidad de contar con un sistema de gestión estándar que facilite la traducción del problema a tareas específicas como: transformaciones de datos, técnicas o algoritmos de análisis de datos, medidas de evaluación de desempeño, así como, la documentación de la experiencia del proceso completo [4]. De esta manera surgen una serie de metodologías orientadas a facilitar la gestión de este tipo de proyectos que tienen como propósito la transformación de datos en conocimiento a través de procesos iterativos conocido como proceso de minería de datos [5].

Una de las metodologías más usuales para el desarrollo de este tipo de proyectos es CRISP-DM (Cross Industry Standard Process for Data Mining) la cual provee un enfoque integral de MD y un modelo de proceso que proporciona un plan completo para llevar a cabo proyectos de MD la cual divide el ciclo de vida del proyecto en seis fases cómo se puede ver en la Figura 1 [6], a continuación se explican las diferentes etapas:

1. Comprensión empresarial: Determinación de los objetivos empresariales, evaluación de la situación actual, establecimiento de objetivos de MD y el desarrollo de un plan de proyecto.
2. Comprensión de datos: Recopilación de datos, análisis exploratorio, identificación de problemas de calidad.
3. Preparación de datos: Selección de variables, corrección, reformateo y transformación de datos.
4. Modelado: Selección de modelo, implementación y obtención de resultados.
5. Evaluación: Evaluación de resultados del modelo.
6. Despliegue: Elaboración del informe e implementación de soluciones.

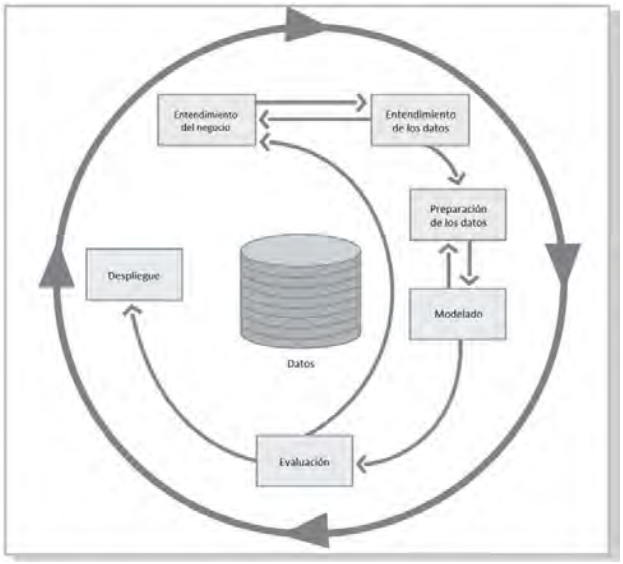



**Figura 1.** Fases del modelo CRISP-DM [7] (Elaboración propia).

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

Entre muchos otros ejemplos de implementación de esta metodología uno de los más recientes se presentan en el trabajo realizado por [8], en el cual se llevó a cabo una estimación de los precios de acciones de mercado en donde se utiliza CRISP-DM como la estructura procedimental orientada a alinear todos los esfuerzos que permitieron asegurar el éxito del problema abordado, explicando a detalle cada uno de los pasos que fueron realizados en cada una de las etapas.

## Entendimiento del problema

La Cadena Productiva de la Uva de Mesa Sonorense (CPUMS) es una de las cadenas productivas agroindustriales más importantes para el estado de Sonora, un ejemplo de esto es que en el año 2020 se tuvo un volumen de producción de 320,990.10 toneladas, representando el 84.47 % a nivel nacional. El 72.6 % de la producción sonorense es exportada a Estados Unidos generando divisas por 733.9 millones de dólares colocándose como el producto agrícola con mayor valor de exportación para sonora [9].

El mercado de Estados unidos presenta diversas oportunidades para la CPUMS como precios altos y que durante 15 o 20 días de comercialización, México es el único proveedor, ya que coincide con la salida de Chile y la entrada de California a este mercado, lo que brinda precios aún más competitivos. Sin embargo, una de las desventajas es que se encuentra caracterizado por un incremento en estándares de calidad, inocuidad y responsabilidad social, provocando un aumento en la estructura de costos para los productores, sumado a precios relativamente estables provoca que los márgenes de utilidad se vean reducidos.

Dado que es un sistema abierto de producción, depende de diversos factores lo que conlleva a que el periodo de comercialización se vuelve incierto, que conjunto a lo mencionado anteriormente la sostenibilidad del sistema se encuentre en riesgo.

Es a partir de estos hechos, que es preciso comprender el comportamiento de los precios en el mercado estadounidense, lo que apoyaría a reducir la incertidumbre y al mismo tiempo permita generar estrategias y tomar decisiones que apoyen a los productores.

## Comprensión de los datos

Para el lograr el objetivo del presente estudio se utilizaron datos disponibles en el repositorio del portal "Agricultural Marketing Service", del Departamento de Agricultura de los Estados Unidos (USDA). La base de datos que se conformó comprende los incluidos desde 1998 al 2019 que forman 880,071 registros con una estructura de incluye 29 variables, las cuales son descritas en la Tabla 1.

Tabla 1. Descripción de los datos (Elaboración propia)

| Variable | Tipo | Descripción | Variable | Tipo | Descripción |
|---|---|---|---|---|---|
| Commodity Name | Texto | Nombre del producto comercializado. | Item Size | Texto | Tamaño del producto. |
| City Name | Texto | Ciudad de la terminal o central de abasto. | Color | Texto | Color del producto. |
| Type | Texto | Orgánica o no. | Environment | Texto | No aplica. |
| Package | Texto | Peso y presentación. | Unit of Sale | Texto | Kilogramos o libras. |
| Variety | Texto | Variedad. | Quality | Texto | Nivel de calidad. |
| Sub Variety | Texto | Subvariedad. | Condition | Texto | Condición del producto. |
| Grade | Texto | Producida en Estados Unidos o fuera de. | Appearance | Texto | Apariencia externa del producto. |
| Date | Texto | Fecha de cotización. | Storage | Texto | No aplica. |
| Low Price | Numero | Cotización mínima alcanzada. | Crop | Texto | Discrimina si el producto es nuevo. |
| High Price | Numero | Cotización máxima alcanzada. | Repack | Texto | Si fue reempaquetada o no. |
| Mostly Low | Numero | Cotización generalmente mínima. | Trans Mode | Texto | Modo de transporte del país de origen. |
| Mostly High | Numero | Cotización generalmente máxima. | Offerings | Texto | Comentarios acerca de la oferta. |
| Origin | Texto | País de origen de la mercancía. | Market Tone | Texto | Estado del mercado de cotizaciones. |
| Origin District | Texto | Distrito de origen de la mercancía . | Price Comment | Texto | Comentarios acerca del precio. |
| Comments | Texto | Comentarios generales. | | | |

Debido a que el objeto de estudio es la uva de mesa sonorense la variable "Origin" que hace referencia al país de origen de la uva se filtró a "México" reduciendo el total de registros de 880,071 a 72,107 referente a la variable "Origin District" que es el estado o región de donde fue enviado y no producido y considerando que sonora es el productor del 84.47 % de la uva de mesa a nivel nacional, se optó por no filtrarla y se consideró que el total de los datos representa los precios de la uva de sonora. En la Tabla 2 se proporcionan las medidas descriptivas básicas para comprender su comportamiento de manera general e identificar posibles problemas de calidad.

Tabla 2. Estadística descriptiva de los datos (Elaboración propia)

| Estadística Descriptiva | Registros | Sin registro | Promedio | Desv. Est. | Min. | Max. |
|---|---|---|---|---|---|---|
| Commodity Name | 72107 | 0 | | | | |
| City Name | 72107 | 0 | | | | |
| Type | 1733 | 70374 | | | | |
| Package | 72055 | 52 | | | | |
| Variety | 72107 | 0 | | | | |
| Sub Variety | 0 | 72107 | | | | |
| Grade | 2361 | 69746 | | | | |
| Date | 72107 | 0 | 40110.42 | 2205.62 | 35801 | 43808 |
| Low Price | 71190 | 917 | 20.34 | 8.88 | 0.25 | 495 |
| High Price | 71190 | 917 | 21.76 | 8.91 | 0.81 | 500 |
| Mostly Low | 17877 | 54230 | 19.35 | 8.94 | 0.25 | 495 |
| Mostly High | 17877 | 54230 | 20.04 | 9.13 | 0.5 | 495 |
| Origin | 72107 | 0 | | | | |
| Origin District | 3247 | 68860 | | | | |
| Item Size | 66940 | 5167 | | | | |
| Color | 107 | 72000 | | | | |
| Environment | 0 | 72107 | | | | |
| Unit of Sale | 10 | 72097 | | | | |
| Quality | 360 | 71747 | | | | |
| Condition | 4878 | 67229 | | | | |
| Appearance | 2373 | 69734 | | | | |
| Storage | 0 | 72107 | | | | |
| Crop | 1163 | 70944 | | | | |
| Repack | 72107 | 0 | | | | |
| Trans Mode | 2550 | 69557 | | | | |
| Offerings | 3028 | 69079 | | | | |
| Market Tone | 61279 | 10828 | | | | |
| Price Comment | 15210 | 56897 | | | | |
| Comments | 1787 | 70320 | | | | |

En la Tabla 1 y 2 se observa como algunas de las variables tienen pocos o nulos registros haciéndolas irrelevantes para el estudio, así mismo, la variable "Package" que hace referencia al peso, se encuentra en formato de texto imposibilitando realizar operaciones matemáticas, también, las variables de "low y high price" cuentan con un valor mínimo de .25 y .81 respectivamente y con un máximo que supera los 495 en ambas variables y al contar con un valor promedio de 20, puede significar que se cuenta con datos atípicos o erróneos lo que puede generar un sesgo en el estudio, para comprobarlo se realizó una gráfica de dispersión la cual se puede ver en la Figura 2 en donde los círculos en rojo señalan los posibles valores atípicos.

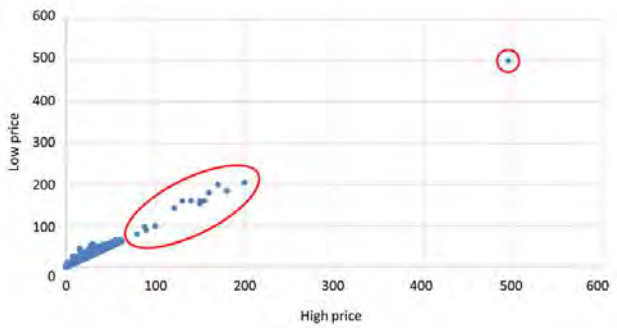



**Figura 2.** Gráfico de dispersión de Low y High Price (Elaboración propia).

Dentro de la base de datos no contaba con una variable que represente el precio por peso de cada una de las transacciones, también se identificó que existe más de un registro por día debido a los diferentes tipos de uva y puntos de envió, que para fines del estudio serán indiferentes y se tomara como un solo día. La variable "Date" que se refiere a la fecha de la transacciona contaba con diferentes formatos como día/mes/año, mes/día/año o año/mes/día imposibilitando trabajar con dichas fechas ya que debe de estar en uno solo.

## Preparación de los datos

En esta etapa se abordan los posibles problemas de calidad encontrados anteriormente y se realiza el procesamiento necesario para que los datos se encuentre listos para ser modelados.

Primero se realizó la selección de las variables de interés para el análisis, siendo: "Package", "Date", "Low y High Price", también se estandarizaron las fechas al formato año/mes/día, permitiendo así gráficas y analizar los datos como una serie de tiempo, proceso crucial para esta etapa.

La variable de texto "Package" requería ser numérica y estandarizar todos los registros a libras, ya que algunos se encontraban en kilos, para esto se elaboró una matriz con los equivalentes y a través de una función de búsqueda se obtuvieron los equivalentes, también se eliminaron aquellos datos faltantes para "Package", "Low y high price" ya que se realizaría el promedio de estas dos últimas variables y posteriormente dividir sobre su peso generando la variable "Precio por libra", en donde en la Figura 3 se puede ver su comportamiento con una gráfica de caja y bigote.

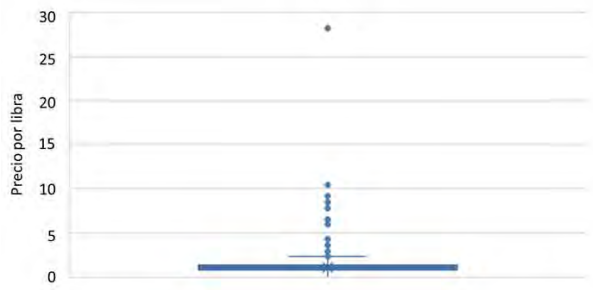



**Figura 3.** Comportamiento del precio de la uva de mesa por año (Elaboración propia).

Según la figura 3 la base de datos cuenta con valores atípicos, los cuales analizándolos de manera individual se identificó que si fueran pesos mexicanos y posteriormente convertidos a dólares, se encontrarían dentro del rango de valores esperados, sin embargo, en el caso de los valores mínimos se encuentran precios de US$ .03 por libra siendo un dato atípico ya que en temporadas de precios bajos tiende a rondar los US$ .40, con el fin de identificar los valores atipicos, se utilizó la media + 3 desviaciones estándar para aquellos valores que tendrían que ser convertidos a dólares y la media − 1.5 desviaciones estándar para eliminar los valores atípicos pequeños. Para convertir los valores a dólares se utilizó una base de datos con el tipo de cambio pesos mexicano a dólar estadounidense del banco de México. Una vez los valores ya estandarizados se procedió a promediar los registros por día e incluir los días faltantes, detalle que condujo al problema de que se encontraban grandes cantidades de días faltantes (74 %) debido a que la uva de mesa se cultiva y comercializa por temporadas específica del año y que el mercado de cotizaciones se encuentra cerrado los días sábados, domingos y festivos, por lo que se optó por utilizar frecuencia mensual y posteriormente imputar los valores a través de regresión lineal, con el fin de no contar con valores faltantes en la base de datos mejorando los resultados obtenidos, el comportamiento de la serie de tiempo se puede ver en la Figura 4.

## Modelado

Dentro de los modelos más utilizados para el análisis de series de tiempo son los modelos autorregresivos integrados de promedios móviles o ARIMA (acrónimo del inglés autoregressive integrated moving average) que tratan de

obtener la representación de los datos en términos de la interrelación temporal de sus elementos [10].

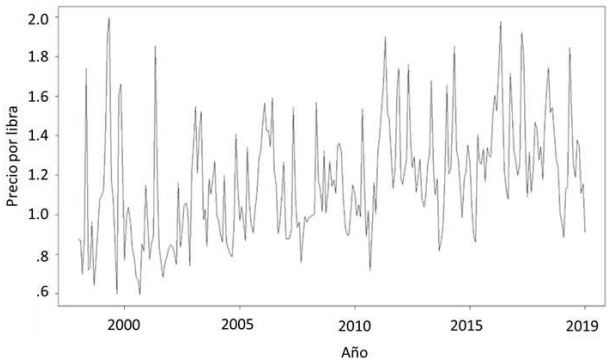



**Figura 4.** Precios mensual de la uva de mesa mexicana en Estados Unidos (Elaboración propia).

En esta etapa se utilizará ARIMA para entender el comportamiento de los precios del año 2019, sin embargo, este modelo requiere que los datos sean estacionarios es decir con una media 0 a lo largo del tiempo y una desviación estándar de 1, en donde en este caso no lo es, por lo que se debe utilizar la diferenciación para lograrlo, también, se deben de identificar los términos autorregresivos (AR) y de Medias móviles (MA) usualmente a través de correlograma y correlograma parcial, siendo un proceso que depende en gran medida de la experiencia y conocimiento de la persona para identificar estos terminos, sin embargo, el lenguaje de programación R brinda una serie de paquetes que automatizan este proceso, mismos que fueron utilizados.

Primero se segmento la base de datos en entrenamiento y prueba para validar la precisión del modelo, en este caso como se analizó el comportamiento de los precios del año 2019, los datos referentes a este año se utilizaron como prueba y del año 1998 a 2018 como entrenamiento.

La función que se utilizó en R para la implementación del modelo ARIMA fue "Auto.Arima" que se encargó de identificar los parámetros óptimos de manera automatizada, que para el conjunto de datos de entrenamiento el modelo óptimo es (1,1,0), a partir del cual se realizó un pronóstico a futuro de 12 meses y otro considerando 1 mes a futuro, en el cual este mes pronosticado fue considerado en el modelo para predecir el siguiente y así hasta que se pronosticaran 12 meses en total, resultados que se pueden ver en la Figura 5.

## Evaluación

Para evaluar la precisión de los modelos se utilizó como referencia la medida de Error Porcentual Absoluto Medio (MAPE) que mide el tamaño de error absoluto en términos porcentuales del pronóstico contra el valor real. Dicho indicador de error se obtuvo utilizando la función "Accuracy" de R la cual como resultados nos indica que el método de 12 meses a futuro tiene un error promedio de 17.22 % mientras que considerando 1 mes a futuro es de 13.09 %, lo que indica que este último es el que brindo mejores resultados.

En la Figura 5 se puede observar como en ambos casos se logra captar la forma general del comportamiento de los precios en donde es evidente que el que utiliza datos más recientes capta de manera más precisa el comportamiento, mientras que el de 12 meses a futuro solo logra captar un comportamiento general.

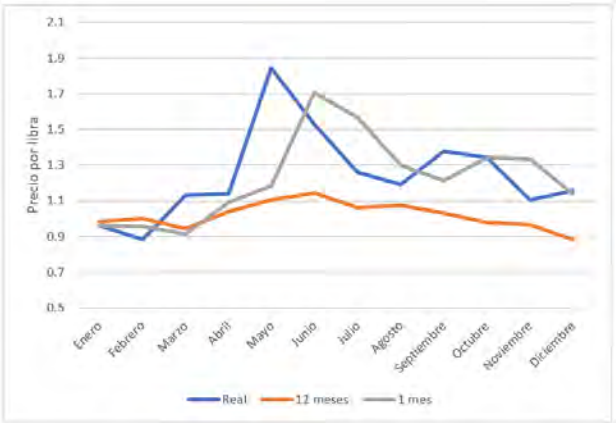



**Figura 5.**Gráfico de valores pronosticados (Elaboración propia).

## Despliegue

A través de este modelo de pronóstico los productores mexicanos de la uva de mesa tendrán una herramienta que les pudiera permitir identificar cuáles podrían ser los posibles precios de comercialización del año siguiente y a su vez utilizarla para identificar aquellos periodos de tiempo donde se obtienen los mejores precios y planear su producción con el fin de incrementar los márgenes de utilidad.

## Conclusiones

La implementación de la metodología CRISP-DM, permitió el desarrollo ordenado del proceso de análisis de datos, ya que a través de ésta se realizó con congruencia estructural el proyecto planteado, así como el logro de su objetivo. El tomar en cuenta cada una de las etapas críticas desde la definición del problema, la gestión de datos, el modelado y la validación de resultados otorga a esta metodología un alto grado de flexibilidad y replicabilidad, que permite el abordaje de una alta gama de distintos tipos de situaciones a analizar. El uso de los modelos ARIMA para el análisis de series de tiempo, permitió una mejor comprensión del comportamiento de los precios de la uva de mesa sonorense en el mercado

estadounidense, lo que proporcionó información y conocimiento vital para dar soporte al proceso de toma de decisiones orientada a la planeación del proceso de producción, así como en la definición de las estrategias de comercialización.

## Trabajos futuros

Para trabajos futuros se podría utilizar una base de datos con mayor cantidad de registros para poder utilizar la frecuencia de manera diaria, también se recomienda utilizar métodos de pronostico mas robustos para lograr capturar de manera más precisa el comportamiento de los precios.✷

## REFERENCIAS


1. Kantardzic, M. (2011). Data Mining: Concepts, Models, Methods, and Algorithms: Second Edition. Segunda edición, 1-550.

2. Fayyad, U., Piatetsky-Shapiro, G., y Smyth, P. (1996). Knowledge Discovery and Data Mining: Towards a Unifying Frame Work. *AAAI Press*, 96, 82–88.

3. Ho., A. (2017). Beyond the Dataset: Understanding Sociotechnical Aspects of the Knowledge Discovery Process Among Modern Data Professionals. *UWSpace*.

4. Wirth, R., y Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. En *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, London, UK: Springer-Verlag, 2000, 29–39.

5. Kotu, V., y Deshpande, B. (2015). Data Mining Process. *Elsevier Predictive Analytics and Data Mining*, 1, 17–36.

6. Shearer, C. (2000). The CRISP-DM model: The New Blueprint for Data Mining. *J. Data Warehous.*, 5(4), 13–22.

7. Pete, C., Julian, C, Randy, K., Thomas, K., Thomas, R., Colin, S., y Rüdiger, S. (2000). Crisp-Dm 1.0. *CRISP-DM Consortium*, 76.

8. Gupta, A., Bhatia, P., Kashyap, D., y Pritesh, J. (2019. Stock Market Prediction Using Data Mining Techniques. En *2nd International Conference on Advances in Science & Technology (ICAST) 2019* por K J Somaiya Institute of Engineering & Information Technology, Mumbai, India, 5.

9. Gobierno del Estado de Sonora. (2020). Estadística uva de mesa sonorense. Secretaría de Agricultura y Desarrollo Rural, recuperado el 24 de mayo de 2021, `http://oiapes.sagarhpa.sonora.gob.mx/portada2.htm`.

10. González, M. (2009). Análisis de series temporales: Modelos ARIMA. Universidad del País Vasco, 169.


## SOBRE LOS AUTORES

**Vicente Solis Sandoval** es egresado de la Universidad de Sonora en la carrera de Ingeniería Industrial y Sistemas (2018). Actualmente estudia el posgrado en Ingeniería en Sistemas y tecnología en la misma Universidad para obtener el título de maestro. Sus áreas de interés son el análisis de series de tiempo, inteligencia artificial y gestión del conocimiento.

**Juan Martin Preciado Rodriguez** Doctorado en Ingeniería en el área de Computación por el Instituto de Ingeniería de Universidad Autónoma de Baja California. Actualmente adscrito a la Coordinación de Desarrollo Regional del CIAD, donde es responsable de los cursos de Métodos y Técnicas de Análisis de Datos I y II, impartidos en los programas de Maestría y de Doctorado en Desarrollo Regional. Las líneas de investigación son el desarrollo y aplicación de algoritmos de extracción de conocimiento en bases de datos.

**Luis F. Romero-Dessens** es Ingeniero Industrial y Maestro en Administración por la Universidad de Sonora, y Doctor en Ciencias de la Ingeniería por la Universidad de Warwick. Es autor en diversas publicaciones académicas en el campo de la ingeniería y de la administración. También ha dirigido tesis y trabajos profesionales en licenciatura, maestría y doctorado. Es miembro de la Asociación Internacional de Ingenieros (IAEng). Sus áreas de interés son diseño, modelación, optimización y simulación en cadenas de suministro (Supply Chains).

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

# IA & Educación

**Lucía Barrón**, **Julieta Noguez** y **Yasmín Hernández**
iaeducacion@komputersapiens.org

## Tecnología educativa

En el ámbito educativo el uso de la tecnología es cotidiano desde hace varias décadas, y hoy en día es común encontrar numerosos recursos didácticos que apoyan el aprendizaje, no solo en el ambiente escolar o laboral sino también en ambientes informales. Los espacios educativos se han transformado y el antiguo salón de clases se ha convertido en un ecosistema donde conviven maestros, estudiantes, dispositivos y sistemas de software con el objetivo de propiciar el aprendizaje. Por ejemplo, los sistemas de realidad virtual y realidad aumentada proporcionan a los estudiantes una experiencia lúdica al estar inmersos en los contextos de aprendizaje, ver Figura 1.

**Figura 1.** Realidad virtual en la educación.

Muchos términos se utilizan de forma indistinta para referirse al uso de la tecnología en la educación o instrucción; por ejemplo, tecnología educativa, tecnología instruccional, diseño instruccional, ingeniería educativa, etc. En Estados Unidos la Asociación para Comunicación y Tecnología Educativa (AECT por sus siglas en inglés) es una asociación académica y profesional que promueve los usos educativos de la tecnología. Ely [1] comenta que desde 1963 la AECT ha revisado y actualizado las definiciones de los conceptos que se utilizan en esta área generando seis publicaciones en diferentes tiempos, así que el concepto de tecnología educativa ha evolucionado y se ha actualizado.

Según Richey [2], la AECT define la tecnología educativa como "*el estudio y la práctica ética de facilitar el aprendizaje y mejorar el desempeño mediante la creación, el uso y la gestión de procesos y recursos tecnológicos adecuados*". Esta definición enfatiza la idea de facilitar el aprendizaje y mejorar el desempeño del aprendiz en un ambiente educativo a través de la creación de recursos tecnológicos. Por su parte, Reiser prefiere utilizar el término Diseño Instruccional y Tecnología (IDT por sus siglas en inglés) porque considera que el término tecnología educativa generalmente se relaciona con el uso de medios con propósito de instrucción y el IDT comprende mucho más que eso. Para Reiser [3], el IDT "*involucra el diseño sistemático de soluciones para problemas en el aprendizaje y el desempeño en el lugar de trabajo; situaciones en las que muchos miembros de una audiencia objetivo no están aprendiendo adecuadamente y/o no están desempeñando su trabajo de manera satisfactoria*".

El proceso sistemático de diseño instruccional y tecnología se basa en el uso de modelos de diseño instruccional [4]. El proceso para efectuar un diseño instruccional considera varias fases: determinar las necesidades del aprendiz, definir las metas y objetivos de la instrucción, diseñar y planear recursos, tareas y evaluaciones y diseñar actividades de enseñanza aprendizaje para asegurar la calidad de la instrucción.

Por otro lado, la inteligencia artificial (IA) es una rama de la computación que a través de sus métodos y técnicas permite la creación de sistemas inteligentes que son capaces de realizar tareas que anteriormente eran exclusivas para los seres humanos. En el área de educación, los sistemas enseñanza-aprendizaje inteligentes, a través de técnicas de la inteligencia artificial, permiten generar experiencias de aprendizaje en las cuales se usa como modelo el razonamiento humano. Este tipo de programas poseen un algoritmo que puede resolver problemas, formular diagnósticos y provee explicaciones a los estudiantes [5].

Se han desarrollado sistemas inteligentes para asistir a estudiantes en diversas tareas administrativas, pero también para apoyar su aprendizaje de acuerdo con sus necesidades particulares. Algunos ejemplos de estos sistemas son el servicio de asesoría virtual para estudiantes que fue creado en la Universidad Deakin de Australia usando IBM Watson [6], el cual responde preguntas básicas de los estudiantes. Otro ejemplo es la plataforma web móvil de aprendizaje adaptativo llamada Osmosis que fue creada para revolucionar la forma en que los estudiantes de medicina aprenden, cuenta con más de 40,000 usuarios a nivel mundial y sus creadores afirman que el sistema maximiza el aprendizaje y la retención de los estudiantes a través del uso de conceptos educativos basados en evidencia, tales como preguntas, tarjetas informativas, videos, imágenes correlacionadas con anclajes de memoria, repetición adaptativa espaciada, aprendizaje colaborativo y gamificación [7].

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) organiza desde hace 8 años la Semana del Aprendizaje Móvil (MLW por sus siglas en inglés) [8], que, en su edición de octubre de 2020, con la pandemia COVID-19 como fondo, llevó como título "Inteligencia Artificial e Inclusión" y abordó el tema de "Más allá de la perturbación: Futuros del aprendizaje habilitados por la Tecnología". El evento generó un "Compendio de iniciativas prometedoras" y ofrece una plataforma para demostrar aplicaciones y prácticas que aprovechan la inteligencia artificial para promover la inclusión y la equidad en la educación, para crear sociedades del conocimiento inclusivas. El desarrollo de tecnología educativa integrando inteligencia artificial permite construir productos que contribuyen a mejorar el aprendizaje. Reiser [3] identifica algunas tendencias y retos en el Diseño Instruccional y Tecnología, las cuales se listan a continuación.

**Mejorar el rendimiento:** se deben diseñar intervenciones para permitir a los individuos mejorar su capacidad para desarrollar las habilidades que necesitarán emplear en sus trabajos, para lograrlo, es necesario que, durante el proceso de aprendizaje, las tareas sean lo más apegadas a las que se espera se realicen en el ambiente de trabajo.

**Apoyo al rendimiento:** proveer al usuario las herramientas e información necesarias para realizar una tarea en el momento que lo requiera, e. g. un sistema GPS para trasladarse de un lugar a otro. Estos elementos de apoyo no siempre están disponibles en los sistemas y representan un reto para los diseñadores instruccionales.

**Aprendizaje en línea:** representa una oportunidad para los diseñadores porque se ha incrementado el interés de las instituciones en esta modalidad y también existe la necesidad de minimizar las deserciones de estudiantes en programas en línea. El número de personas atendiendo cursos en línea creció en los últimos 10 años de ser en 2010 menor a 10

**Redes sociales:** han transformado la comunicación entre individuos provocando que los recursos que se generan sean compartidos sin necesidad de un ambiente educativo. Esto representa un reto para los diseñadores instruccionales que deben decidir cómo identificar, incorporar y administrar recursos educativos entre todas las oportunidades de aprendizaje expuestas para los aprendices.

**Juegos educativos:** diseñados explícitamente para lograr el aprendizaje de elementos específicos, se usan comúnmente en niveles de educación básica (K-12). Transferir habilidades necesarias en las tareas del mundo real a partir de un juego educativo es un reto para los diseñadores instruccionales.

**Aprendizaje móvil:** los dispositivos móviles como teléfonos inteligentes, tabletas y lectores electrónicos se utilizan frecuentemente en actividades de aprendizaje, pero son también fuente de distracción en estos ambientes. El reto al usar estos dispositivos es reducir posibles distractores.

**Cursos Abiertos Masivos en Línea:** son cursos conocidos como MOOCs frecuentemente ofertados por universidades y donde participa un gran número de estudiantes dispersos geográficamente. La alta tasa de estudiantes que abandonan los MOOCs representa un reto para los diseñadores que deben tratar de responder la pregunta ¿qué clase de actividades instruccionales y herramientas de apoyo para estudiantes deben incluirse en estos cursos?

**Analíticas de aprendizaje:** se define como "la colección y análisis de datos sobre estudiantes con el propósito de entender y optimizar su aprendizaje" [3]. El reto para el diseñador instruccional es identificar cuáles datos deben ser recolectados y cómo deben estos ser usados para mejorar el aprendizaje.

Considerando los diversos escenarios (escolar o laboral) donde se utilizan recursos educativos digitales, queda claro que el desarrollo de estos elementos no es una tarea trivial y es necesario contar con personas especializadas para realizar este trabajo.

La Tecnología Educativa que se incorpore a ambientes educativos debe cumplir con varios objetivos como: facilitar el aprendizaje y mejorar el desempeño de los estudiantes, así como promover la inclusión y la equidad en la educación; para cumplir estos objetivos es indispensable el uso de la inteligencia artificial a través de la cual se incorporen componentes innovadores en los procesos de enseñanza-aprendizaje para beneficiar a los estudiantes.✳

## REFERENCIAS


1. Ely, D.P. (2008). Reflections on the 2008 AECT Definitions of the Field. *TechTrends*, 52, 25.

2. Richey, R.C. (2008). Reflections on the 2008 AECT Definitions of the Field. *TechTrends*, 52, 24.

3. Reiser R.A. (2017). Eight Trends Affecting the Field of Instructional Design and Technology: Opportunities and Challenges. In: Lai FQ., Lehman J. (eds) Learning and Knowledge Analytics in Open Education. Springer, Cham.

4. Dick, W., Carey, L., y Carey, J. O. (2014). Systematic design of instruction (8th ed.). Boston, MA: Pearson.

5. Rodríguez-Chávez M. H. (2021). Sistemas de tutoría inteligente y su aplicación en la educación superior. *RIDE. Revista Iberoamericana de Investigación y Desarrollo Educativo*, 11(22)

6. IBM Watson for Students and Education. `http://virtuallyinspired.org/portfolio/deakin-university/`.

7. Osmosis Adaptive Learning Tool for Medical School. `http://virtuallyinspired.org/portfolio/osmosis/`.

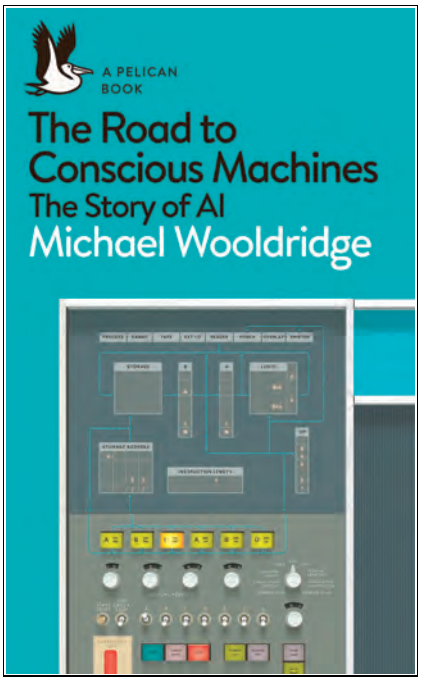
8. UNESCO, Artificial intelligence and inclusion, compendium of promising initiatives: Mobile Learning Week 2020. `https://unesdoc.unesco.org/ark:/48223/pf0000374644`.

# Deskubriendo Konocimiento

**Alejandro Guerra Hernández** y **Leonardo Garrido**
deskubriendokonocimiento@komputersapiens.org

## Revisión de "The Road to Conscious Machines: The Story of AI" *

**Dr. Alejandro Guerra Hernández**
Universidad Veracruzana
Instituto de Investigaciones en Inteligencia Artificial




**Portada de la revista.***

Otra historia de la Inteligencia Artificial (IA). ¿Realmente el mundo necesitaba otro libro de divulgación de la IA y sus orígenes? Así inicia este libro de Michael Wooldridge, jefe del Departamento de Ciencias de la Computación de la Universidad de Oxford. Es que tengo un nuevo ángulo para contarla– bromea Wooldridge: ¡Es la historia de la IA a través de sus fallos! Curiosa aproximación de quién recibió este Julio el reconocimiento como investigador Turing, líder-mundial en IA, Programa de posicionamiento del Reino Unido para el avance y éxito de esta área del conocimiento.

El título del libro es engañoso, no se trata exclusivamente de una historia de la IA, ni mucho menos de una hoja de ruta para la creación de máquinas conscientes. En realidad el texto persigue tres objetivos. El primero es contar lo que es la IA y, de manera muy importante, tratar de dejar en claro lo que no es. Para ello el autor diferenciará lo que llama el gran sueño: Construir máquinas que tienen las mismas capacidades que nosotros para actuar de forma inteligente; de lo que hacemos realmente en el día a día: Lograr que las máquinas hagan tareas específicas que requieren cerebros humanos y, potencialmente cuerpos humanos, para las que la computación tradicional no provee una solución. La traducción automática, los automóviles autónomos, la detección de tumores en rayos-x y ultrasonidos, son ejemplos de estas tareas. Para entender la divergencia entre lo que hacemos realmente y el gran sueño, Wooldridge se adentrará en las razones por las cuales es muy difícil crear IA, vía la explicación de lo que las computadoras pueden hacer y lo que no. Esta parte del libro consta de un sólo capítulo, "Cerebros electrónicos de Turing", que ubica el nacimiento de la IA y la computación, en el King's College, año 1935.

La segunda parte del libro tiene como objetivo contar propiamente la prometida historia de la IA. Wooldridge explota los siete argumentos básicos propuestos por Christopher Brooker (The Seven Basic Plots, Bloomsbury, 2004) para decidir cómo contarla, más allá del orden cronológico de los eventos. Aunque, argumenta el autor, muchos de nuestros colegas se decantarían por una historia típica de pobres a ricos, del estilo el éxito comenzó en un obscuro garage... él preferiría una historia del tipo vencer al monstruo, donde el enemigo es la complejidad computacional; o del tipo búsqueda épica al estilo del Señor de los Anillos. Pero finalmente se decanta por un argumento del tipo caída y resurgimiento, aunque aclara, la historia de la IA es tristemente como la de Sísifo – un interminable renacer para volver a caer. De ahí el ángulo adoptado por Wooldridge para esta historia de la IA, que persigue aportar, en sus propias palabras, una necesaria gran dosis de humildad **.

*Wooldridge, Michael. The Road to Conscious Machines: The Story of AI. Pelican, United Kingdom, 2021.

**Quien crea que el triunfalismo ramplón es cosa del pasado, debería leer el reporte de Will Douglas Heaven sobre las aportaciones de la IA ante el COVID, en el MIT Technology Review: `https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/`

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

Este es pues el tema de la segunda parte del libro, organizada en cuatro capítulos que narran las grandes caídas y renacimientos de la IA. "La edad de oro" abarca el periodo que va de la escuela de verano en Dartmouth College, convocada por John McCarthy en 1956, al invierno de los años 1970s que siguió a las duras críticas de Dreyfus en Estados Unidos (Alchemy and AI, 1965) y el reporte Lighthill en Reino Unido (1972). El periodo está caracterizado por la especialización de la IA en sus conocidas subdisciplinas: visión, aprendizaje, lenguaje natural, etc., y por el desmedido entusiasmo ante los rápidos avances logrados por las técnicas de resolución de problemas y búsqueda que, como señalaron los críticos mencionados, habían minimizado el problema de la explosión combinatoria.

"Conocimiento es Poder" aborda el reinado de los Sistemas Expertos a finales de los años 1970s y principios de los 80s. La idea central aquí es que hacía falta explotar el conocimiento de los humanos para hacer avanzar la IA. Las diferentes técnicas de representación de conocimiento son producto de esta época, incluyendo los Sistemas Basados en Reglas, la Programación Lógica, y las Ontologías. Desafortunadamente, el sentido común, el razonamiento bajo contradicciones y otros problemas resultaron muy difíciles de resolver y, aunado esto a las dificultades encontradas para extraer y validar conocimiento, condujeron a una perdida de interés.

"Robots y Racionalidad" aborda la revolución de la llamada IA basada en el comportamiento en los años 1990s y su énfasis en las formas robóticas y la interacción con el medio ambiente. En esta época nace el interés en los agentes inteligentes. De especial interés es el surgimiento de un nuevo paradigma para la IA: La construcción de agentes capaces de actuar en nuestro beneficio, la idea central del tan celebrado libro de Peter Russell y Stuart Norvig (Artificial Intelligence a Modern Approach, 4th ed., Pearson, 2020); y la adopción de formalismos como la teoría de la decisión y el razonamiento probabilista.

"Avances profundos" inicia en 2014 con la compra de DeepMind que inaugura la revolución de la IA basada en el aprendizaje automático y las redes neuronales. En realidad esta tradición tiene una larga historia en la IA que inicia con el trabajo de McCulloch y Pitts en 1943 y el perceptrón de Rosenblath en los años 1950s. El renacer de las redes neuronales está marcado por el libro Parallel Distributed Processing (1985) y el algoritmo de retropropagación. Y en un tercer momento, el surgimiento del aprendizaje profundo. El capítulo señala las conocidas limitaciones de esta tecnología: su falta de transparencia y las sutiles y extrañas maneras que tienen de fallar en su cometido. También resalta sus grandes logros: La auto-programación desde cero ejemplificada con los juegos de Atari y, evidentemente, el triunfo en Go. Para finalizar esta parte, Wooldridge nos advierte de un aparente cisma entre la comunidad de Aprendizaje Automático y la de IA, y explica sus orígenes en la historia que aquí termina.

La tercera parte tiene como objetivo contar que es lo que los sistemas de IA pueden hacer actualmente y qué podrían hacer pronto, así como sus limitaciones. Esta parte está compuesta por cuatro capítulos. "La IA hoy" reconoce los avances científicos genuinos que, combinados con la disponibilidad de grandes volúmenes de datos y poder de cómputo muy barato, han hecho posibles sistemas de IA que parecerían milagrosos a los ojos de los fundadores del área. Eso se ilustra con las aplicaciones al cuidado de la salud y la producción de autos autónomos. Los dos capítulos restantes tienen que ver con los miedos que la IA despierta. "Como imaginamos que las cosas pueden ir mal", aborda los temores infundados, derivados en gran medida del por ahora remoto gran sueño de la IA: La Singularidad y las narrativas estilo Terminator; la perdida del control sobre nuestros sistemas; y los falsos dilemas morales. "Como las cosas pueden ir realmente mal" aborda los temores justificados: el impacto laboral, las armas autónomas, el sesgo algorítmico, las noticias falsas y la falsa IA (esos sistemas que parecen IA pero no lo son). El último capítulo "¿Máquinas conscientes?" retoma el gran sueño y por tanto es más especulativo que el resto del libro. Dado que el autor del libro y quién esto escribe, somos expertos en Sistemas Multi-Agentes, resalto la observación de que necesitamos considerar los aspectos sociales de la inteligencia si queremos avanzar hacía el gran sueño.

Wooldridge nos advierte que el libro es muy selectivo dada su naturaleza. Totalmente de acuerdo, yo eché de menos las aportaciones de Allen Newell y su Teoría Unificada de Cognición (Harvard University Press, 1990). Aún así, creo que el libro cumple cabalmente con su cometido: Darnos una idea de qué es la IA y hacía donde se dirige. Muy recomendable como una lectura complementaria en los cursos de introducción a la IA.✳

© 2021 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

# Si te cuidas **tú**, nos cuidamos **todos**

Todos podemos hacer algo para mitigar y contener la expansión de las enfermedades. Lo primero siempre es entender qué son, cómo se transmiten y cómo afectan.

En este caso, se trata del coronavirus SARS-COV2. Apareció en China en diciembre pasado y provoca una enfermedad llamada COVID-19, que se ha extendido por el mundo y fue declarada pandemia global por la Organización Mundial de la Salud.

## Medidas de prevención

LAVA TUS MANOS CON AGUA Y JABÓN

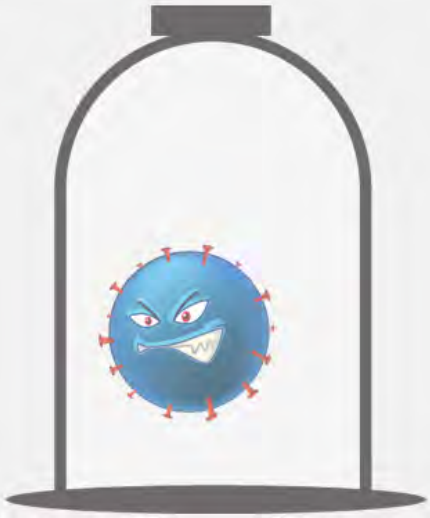
ACUDE AL MÉDICO SOLO EN UNA URGENCIA

TOSE O ESTORNUDA EN LA PARTE INTERNA DE TU CODO

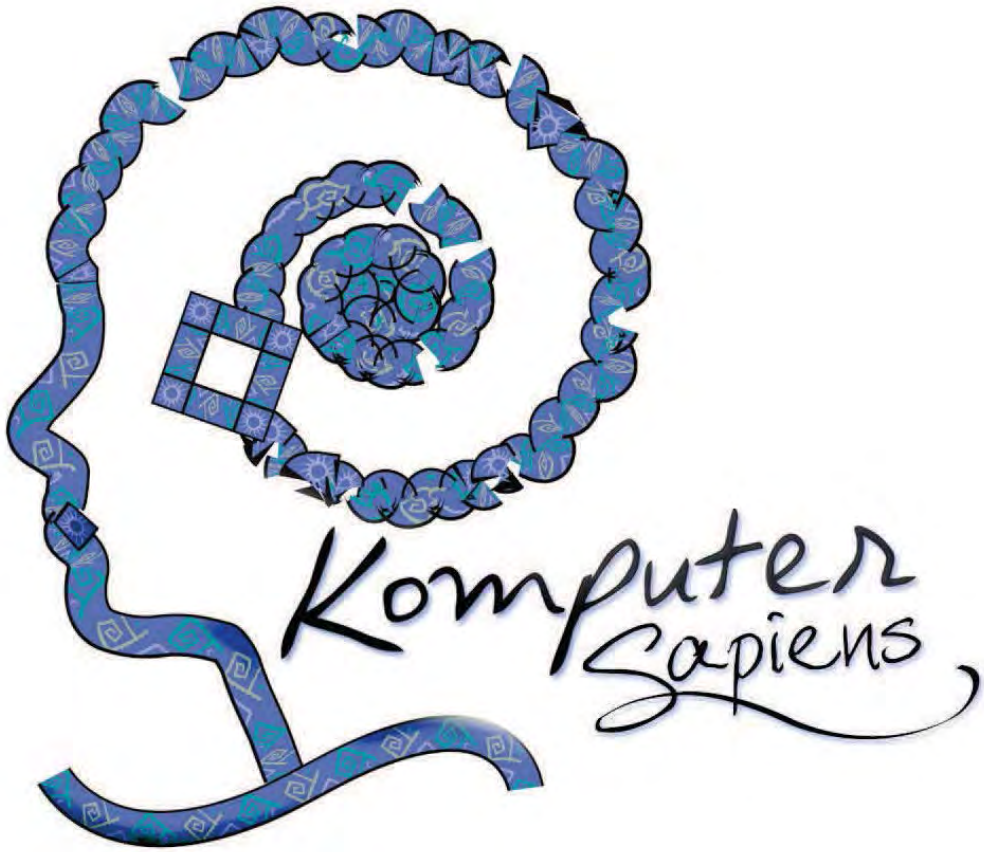
NO DIFUNDAS NOTICIAS FALSAS

Atendiendo las normas de la Nueva Normalidad y cuidándonos unos a otros podremos salir adelante.

**¡SI AISLAMOS AL VIRUS, LO VENCEREMOS!**



Información tomada de: https://coronavirus.gob.mx




## ¡Publique en Komputer Sapiens!

CONACYT
ÍNDICE DE REVISTAS MEXICANAS DE DIVULGACIÓN CIENTÍFICA Y TECNOLÓGICA (IRMDCT)

**Komputer Sapiens** solicita artículos de divulgación en todos los temas de Inteligencia Artificial, dirigidos a un amplio público conformado por estudiantes, académicos, empresarios, tomadores de decisiones y consultores. Komputer Sapiens es patrocinada por la **SMIA, la Sociedad Mexicana de Inteligencia Artificial**

SMIA
www.smia.org.mx

**Instrucciones para autores e información general:** http://www.komputersapiens.org
**Síguenos en las redes sociales:** www.facebook.com/Komputer.Sapiens, twitter.com/KomputerSapiens