www.komputersapiens.org

KOMPUTER SAPIENS

Revista de Divulgación de la Sociedad Mexicana de Inteligencia Artificial

Año 16
Volumen 3
sept-dic 2024



IA para detectar plagas en cultivos de aguacate

Monitoreo inteligente para adultos mayores

Tecnologías de generación de energía limpia con IA

IA en el procesamiento de lenguaje natural

Detectando toxixidad en las plataformas digitales

Visión estereoscópica para medir distancia

Predicción del Movimiento de Peatones

IA en la industria turística

Detección de ataques phishing

Modelado e identificación de robots

Origamis en la Robótica Suave

ISSN 2007-0691

9 772007 069007

©Komputer Sapiens, Año XVI Volumen III, septiembre-diciembre, 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 31 de agosto 2024.
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

Presidenta	María Lourdes Martínez Villaseñor
Vicepresidenta	Hiram Eredín Ponce Espinosa
Secretario	Roberto Antonio Vázquez Espinosa de los Monteros
Tesorero	Gilberto Ochoa Ruiz
Vocales:	Gustavo Arroyo Figueroa
	Noé Alejandro Castro Sánchez
	Hugo Jair Escalante Balderas
	Sofia Natalia Galicia Haro
	Antonio Marín Hernández
	Bella Citlali Martínez Seis
	Iris Iddaly Méndez Gurrola
	Leobardo Morales Tiburcio
	César Jesús Núñez Prado
	Iván Alejandro Olvera Rodríguez
	Obdulia Pichardo Lagunas
	David Eduardo Pinto Avendaño
	Victor Josué Ruiz Martínez

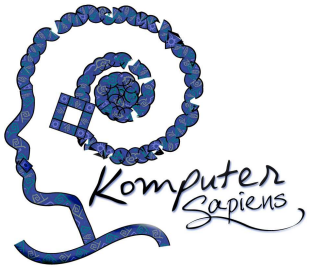
Komputer Sapiens

Directora general	María Lourdes Martínez Villaseñor
Editora en jefe	Karina Mariela Figueroa Mora
Editor científico	Héctor Gabriel Acosta Mesa
Editores asociados	Marco Antonio Aguirre Lam
	Laura Cruz Reyes
Editores invitados	Jezreel Mejía
	Mirna Muñoz
Coordinadora de producción	Viridiana Mena Gómez
e-Tlakuilo	Karina Figueroa
	Rosa María Valdovinos
Estado del IArte	Reyna Carolina Medina Ramírez
	Jorge Rafael Gutiérrez Pulido
Sakbe	Claudia G. Gómez Santillán
	Laura Cruz Reyes
	Laura Gómez Cruz
IA & Educación	J. Julieta Noguez Monroy
	María Lucía Barrón Estrada
	María Yasmín Hernández Pérez
Deskubriendo Konocimiento	Alejandro Guerra Hernández
	Gildardo Sánchez Ante
Corrección de estilo	Claudia Leticia Díaz González
	Gilberto Rivera Zárate
	Guillermo de Jesús Hoyos Rivera
	Leonardo Romero Muñoz
	Marcela Quiroz Castellanos
	Nelson Rangel Valdez
	Nicandro Cruz Ramírez
	Sonia Lilia Mestizo Gutíerrez
	Susana Aurora Velasco Montiel

Directores Fundadores
Carlos Alberto Reyes García
Ángel Kuri Morales

Comité Editorial
Ildar Batyrshin
Oscar Herrera Alcántara
Jesús Favela Vara
Sofía Natalia Galicia Haro
Miguel González Mendoza
Raúl Monroy Borja
Eduardo F. Morales Manzanares
Leonardo Garrido Luna
Carlos Alberto Reyes García
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
Félix Castro

Árbitros
Adriana Iñiguez Carrillo
Adriana Laura López-Lobato
Adriana Peña
Alberto Ochoa-Zezzatti
Aldo Márquez-Grajales
Alejandra Silva
Angel Kuri-Morales
Angel Ortiz
Angel Sánchez García
Antonio Camarena-Ibarrola
Antonio Marin Hernandez
Aurelio Lopez-Lopez
Candy Sosa Jimenez
Carlos Reyes
Carlos Toxtli
Carmen Mezura-Godoy
Christian Casillas
Edgard Benítez-Guerrero
Eduardo Morales
Efrén Juárez Castillo
Ek del Val
Federico Alonso-Pecina
Fernando Arambula
Fernando Montes
Gabriel García-Mireles
Gilberto Ochoa-Ruiz
Guillermo Hoyos-Rivera
Gustavo Vargas Hakim
Hector Acosta-Mesa
Héctor Aviles-Arriaga
Hector Cardona-Reyes
Homero Vladimir Rios Figueroa
Hugo Jair Escalante
Hugo Mitre Hernández
Humberto Sossa
Ivan Garcia
Jaime Cerda
Jorge Rodas-Osollo
Jose Alberto Hernandez-Aguilar
Jose Francisco Martínez Trinidad
José Luis Reyes
Jose Sánchez-Cervantes
Juan Ahuactzin
Julio Ponce
Karina Figueroa Mora
Karina Gutiérrez-Fragoso
Karla Olmos-Sánchez
Leonardo Tapia Figueroa
Lisbeth Rodríguez
Luis Colombo-Mendoza
Luis A. Castro
Luis Fernández-Martínez
Manuel Montes-y-Gomez
María del Pilar Salas Zárate
Marisol Flores-Garrido
Miguel Hidalgo-Reyes
Mirna Munoz
Nancy Pérez-Castro
Rafael Rivera-Lopez
Ramon Brena
Raúl Aguilar Vera
Raúl Monroy Borja
Rene Francisco Navarro Hernandez
Rocío Barrientos-Martínez
Sandra Nava-Munoz
Saul Dominguez-Isidro
Sergio Hernandez Mendez
Sodel Vázquez Reyes
Sonia Mestizo-Gutiérrez
Sussy Bayona Ore
Tania Turrubiates-Lopez
Victor Morales
Víctor Saquicela
Yadira Quiñonez



Contenido

¿A dónde vas?: Aplicaciones de la IA a la Predicción del Movimiento de Peatones
Pág. 8 ⇒Jean Bernard Hayet, David Gómez Gutíerrez, Leobardo Campos Macías y Javier Felip León

Monitoreo inteligente para el cuidado de los adultos mayores en sus hogares
Pág. 14 ⇒ Elizabeth López Lozada, Humberto Sossa Azuela y Elsa Rubio Espino

Cómputo inteligente en el diseño de celdas de hidrógeno para la generación de energía limpia
Pág. 18 ⇒ Luis Moreno Mendieta, Rolando Menchaca Méndez, Adriana Menchaca Méndez y Marisol Flores Garrido

Aplicaciones de software con inteligencia artificial en el procesamiento de lenguaje natural
Pág. 23 ⇒ Sabino Miranda, Obdulia Pichardo-Lagunas y Bella Martínez-Seis

Comparación de dos modelos de redes neuronales preentrenados aplicados en la detección de plagas en cultivos de aguacate
Pág. 28 ⇒ Jesús García Palomera, Jesús Alberto Verduzco Ramírez, Juan García Virgen y Luis Eduardo Morán López

El Arte de los Origamis como Actuadores en la Robótica Suave
Pág. 33 ⇒ Juana-Mariel Dávila-Vilchis, Gustavo Levy Lara Morales, Adriana Herlinda Vilchis González, Melissa Días- Omonte, Juan Carlos Ávila Vilchis y Juan Manuel Jacinto-Villegas

Detección de ataques phishing mediante técnicas de Inteligencia Artificial
Pág. 38 ⇒ Lázaro Bustio-Martínez, Vitali Herrera-Semenets, Jorge A. González Ordiano y Miguel Á. Álvarez Carmona

Develando la detección de toxicidad en las plataformas digitales: El papel de las redes neuronales artificiales
Pág. 44 ⇒ Allan Balam Rueda-Gutíerrez, Alejandro Rodríguez-Molina, Oscar Armando Gómez-Vargas y Eric Hernández-Castillo

Implementación práctica de Matlab y visión estereoscópica para la medición de distancia de una persona
Pág. 50 ⇒ Rodrigo Hinojosa-Cruz, Juan Manuel Jacinto-Villegas y Otniel Portillo-Rodríguez

Revelación de Datos útiles para la Industria Turística de Guanajuato mediante Aprendizaje Profundo y Procesamiento del Lenguaje Natural
Pág. 56 ⇒ Angel Díaz Pacheco, Dora Luz Almanza Ojeda, Ansel Y. Rodríguez González, Miguel Á. Álvarez- Carmona, Ramón Aranda y Mario Alberto Ibarra Manzano

Modelado dinámico/Identificación de robots usando redes neuronales artificiales: un enfoque para principiantes
Pág. 61 ⇒ Jesús A. Medrano-Hermosillo, Abraham E. Rodríguez-Mata, Larbi Djilali, Victor Alejandro Gonzalez Huitron y Oscar J. Suarez Sierra

Columnas

Sapiens Piensa. Editorial pág. 2

e-Tlakuilo pág. 4

Estado del IArte pág. 6

Sakbe pág. 7

IA & Educación pág. 67

Deskubriendo Konocimiento pág. 69

Sapiens Piensa

Jezreel Mejía y Mirna Muñoz

La Inteligencia Artificial (IA) ya es una realidad y esta evolucionando a pasos agigantados, prueba de ello, es como se acerca cada vez más a escenarios de películas como: "El hombre Bicentenario (1999)", "Yo, Robot (2004)", hasta una de las películas mas reciente "El creador (2023)", por mencionar algunas. Por lo tanto, en este número de tema libre se explora la evolución que sigue presentando la IA a través de la selección de 11 artículos que abordan su aplicación en diferentes dominios tales como: la detección de movimientos de los seres humanos, generación de energía limpia, procesamiento de lenguaje natural, detección de plagas, como mejorar actuadores en robótica suave, ciberseguridad, plataformas digitales, visión estereoscópica, sector turístico, y de robots en la industria.

Para iniciar la exploración de esta evolución, el primer artículo **"¿A dónde vas?: Aplicaciones de la IA a la Predicción del Movimiento de Peatones"** escrito por *Bernard Hayet et al.*, presentan como a través de la IA Generativa se puede solucionar el problema de predicción que surge en la navegación autónoma de coches, para inferir los movimientos futuros y las acciones de un humano, con la finalidad de una adopción más amplia de coches autónomos. Además, esta propuesta puede aplicarse a otros dominios de la robótica como la colaboración humano-robot en tareas de manipulación en espacios compartidos.

En el artículo **"Monitoreo inteligente para el cuidado de los adultos mayores en sus hogares"** presentado por *López Lozada et al.*, dan a conocer una propuesta inicial para el seguimiento de personas mayores llamado FairMOT en combinación con el modelo ConvNeXt pre entrenado. Este modelo detecta de 3 a 5 actividades como caerse al suelo, caminar, beber, hablar por teléfono y comer. Los resultados que presentan los autores muestran la efectividad de su propuesta.

Moreno Mendieta et al., a través del artículo **"Cómputo inteligente en el diseño de celdas de hidrógeno para la generación de energía limpia"** nos presentan la urgencia de la búsqueda de alternativas en la generación de energía limpia. Para ello, los autores describen como las redes neuronales artificiales y la computación evolutiva pueden hacer una aportación crucial para mejorar el diseño y acelerar la implementación de celdas de hidrógeno.

El artículo **"Aplicaciones de software con inteligencia artificial en el procesamiento de lenguaje natural"** de *Miranda et al.*, nos presenta una rama de la IA: el procesamiento del lenguaje natural (PLN), la cual se centra en permitir a las máquinas entender, interpretar y generar lenguaje humano de manera coherente. Nos muestran como el uso del PLN, se han desarrollado aplicaciones de software con el objetivo de comprender el contexto de las conversaciones, comprender los sentimientos en el texto, traducir idiomas de manera automática, generar texto de manera creativa y proporcionar respuestas contextuales en interacciones con el usuario.

Por otra parte, *García Palomera et al.*, en el artículo **"Comparación de dos modelos de redes neuronales preentrenados aplicados en la detección de plagas en cultivos de aguacate"** nos presentan la importancia de la detección de plagas en los cultivos de aguacate dada su proyección de producción para el 2023 de 3.16 millones de toneladas. Los resultados que presentan los autores confirmar que los modelos de redes neuronales son efectivos para este tipo de fruta.

En el artículo **"El Arte de los Origamis como Actuadores en la Robótica Suave"** presentado por *Dávila-Vilchis et al.*, analizan el diseño, fabricación y caracterización de los origamis Zigzag, Yoshimura y Kresling con el objetivo de transformarlos en actuadores suaves en sistemas robóticos complejos, por ejemplo en las rotaciones sucesivas de un bisturí, que podría tener como resultado una cirugía mínimamente invasiva.

Bustio-Martínez et al., nos presentan en su artículo **"Los ataques de phishing y la Inteligencia Artificial"** como a través de la IA se puede ayudar en la detección de phishing por medio del a) aprendizaje profundo (Machine Learning) para el análisis de la estructura y contenido de los correos electrónicos y de b) redes neuronales convolucionales (CNN) y redes neuronales recurrentes (RNN) para analizar el contenido visual y temporal de los correos electrónicos.

Rueda-Gutiérrez et al., en el artículo **"Develando la detección de toxicidad en las plataformas digitales: El papel de las redes neuronales artificiales"** los autores abordan la toxicidad de las plataformas digitales, específicamente las redes sociales, ya que son utilizados por usuarios no solo para expresarse sino para proyectar comentarios tóxicos que causan daños a las personas y comunidades. En este sentido las redes neuronales artificiales y sus diferentes arquitecturas son una opción que muestra resultados en la solución del problema de detección de textos tóxicos que permiten prevenir y manejar conductas dañinas en estas plataformas.

Hinojosa-Cruz et al., en el artículo **"Implementación práctica de Matlab®y visión estereoscópica para la medición de distancia de una persona"** presentan una guía práctica, scripts y otros materiales para usar Matlab®y visión estereoscópica en un caso práctico

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

para la detección y medición de distancia de una persona, lo que permite replicar un proceso computacional de imágenes con el fin de detectar objetos, formas, personas y colores dentro de un espacio de trabajo llamado campo de visión.

En el contexto del sector turístico, *Díaz Pacheco et al.*, en su artículo **"Revelación de datos útiles para la industria turística de Guanajuato mediante Aprendizaje Profundo y Procesamiento del Lenguaje Natural"** presentan la aplicación de técnicas avanzadas de procesamiento del lenguaje natural para el análisis de datos provenientes de diversas fuentes, abordando su naturaleza ruidosa y desordenada. Los autores presentan como en Guanajuato lograron detectar las atracciones turísticas más populares como "El Teatro Juárez" y la "Alhóndiga de Granaditas", por mencionar algunos de sus hallazgos.

Finalmente, el artículo 11, presentado por *Medrano-Hermosillo et al.*, **"Modelado dinámico/identificación de robots usando redes neuronales artificiales: un enfoque para principiantes"** describen la importancia del uso de los robots y su amplio crecimiento en la industria. En este contexto, los autores presentan como las redes neuronales artificiales (RNAs) han cobrado importancia para la identificación de sistemas dinámicos, como es el caso de los robots dada sus amplias ventajas respecto a las técnicas tradicionales.

Esperamos que este número permita al lector apreciar una visión general de la constante evolución de la IA hacia diversos dominios. Invitamos a todos los lectores, colegas, estudiantes y personas interesadas, a sumergirse en cada uno de los artículos de su interés.

Queremos agradecer a todos los autores que contribuyeron en el envío de artículos, así como al comité científico por concretar esta edición. ✷

SOBRE LOS AUTORES



Jezreel Mejia es Doctor en Informática por la Universidad Politécnica de Madrid (UPM), España con mención de "Doctorado Europeoz doctorado europeo. Recibió el premio a la mejor tesis doctoral de su generación. En 2019 ha recibido la distinción de un doctorado honoris causa avalado por la secretaria media superior y superior del estado de Morelos. Miembro del Sistema Nacional de Investigadores (SNII) nivel 2. Coordinador de la Maestría en Ingeniería del Software (IS) de CIMAT A.C. Unidad Zacatecas. Ha publicado diversos artículos técnicos en temas relacionados con la Ingeniería del software y de la Inteligencia Artificial hacia la IS, tanto en congresos Nacionales e Internacionales así como, en revistas conocidas como JCRs. Actualmente, es presidente del Congreso CIMPS, fundado en 2012 con una asistencia de 500 personas (autores, instructores, ponentes principales y asistencia en general). `https://www.researchgate.net/profile/Jezreel-Mejia`. Email: `jmejia@cimat.mx`.



Mirna Muñoz es Doctora en Lenguajes y Sistemas Informáticos e Ingeniería de Software por la Facultad de Informática de la Universidad Politécnica de Madrid, con mención de "Doctorado Europeo". Realizó una Estancia posdoctoral en la Universidad Carlos III de Madrid. Actualmente es investigadora titular A del CIMAT- Unidad Zacatecas en el área de Ingeniería de Software, y miembro del SNII nivel 2. Ha publicado 200 artículos técnicos relacionados con la Ingeniería de Software y participa como revisora de varios JCR y como parte de comités de conferencias internacionales. Es vicepresidenta y coorganizadora del Congreso CIMPS y ha sido copresidenta de talleres para la conferencia EuroSPI desde 2018. Recientemente, ha estado liderando la conferencia ENC como coorganizadora del Taller de Ingeniería de Software. Áreas de interés son: Definición y mejora de procesos, factor humano, gestión del cambio, entornos multimodelo, DevOps, Calidad de software, integración de equipos, implementación de modelos y estándares de calidad, métodos ágiles, gamificación y juegos serios. `https://www.researchgate.net/profile/Mirna-Munoz`. Email: `mirna.munoz@cimat.mx`.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

e-Tlakuilo: Cartas de nuestros lectores

Karina Figueroa y 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.

Ajedrez computacional: ¿Cómo la inteligencia artificial se convirtió en la campeona del mundo?

Pregunta: Raúl Salazar Godínez.

Sobre las metodologías específicas de aprendizaje, ¿cómo exactamente AlphaZero implementa el aprendizaje por refuerzo para superar a otros programas avanzados en sólo unas horas?
Sobre la Interacción y retroalimentación humana, el texto menciona que las máquinas han cambiado la forma de interactuar con el ajedrez, pero ¿cómo se están usando esas interacciones para mejorar las propias máquinas (Stock fish, komodo)?

Responder la pregunta sobre las metodologías de aprendizaje por refuerzo que implementa AlphaZero puede hacerse con distintos niveles de profundidad. Una explicación amplia puede ser material para un artículo enfocado en el tema; en está ocasión, se abordará desde una perspectiva más breve. El aprendizaje por refuerzo es una versión mucho más general del proceso de optimización de una función de error, y se integra por cuatro componentes básicos: 1) un agente; 2) un entorno con el que puede interactuar; 3) una serie de acciones o decisiones que puede tomar; y, 4) un método de retroalimentación. Éste último componente refuerza al agente de manera positiva o negativa cuando las decisiones que haga en torno a un objetivo sean buenas o malas, respectivamente. El caso particular de AlphaZero implementa las cuatro componentes de la siguiente manera: a) AlphaZero, quien es el agente; b) el entorno, que es el tablero de ajedrez jugando contra otro agente; c) las acciones, que son los movimientos legales de ajedrez; y, d) la retroalimentación, que se realiza al saber si se ha ganado, perdido o empatado una partida. Para que AlphaZero aprenda de manera exitosa, el agente interactúa con el entorno, obtiene conocimiento a partir de las decisiones que toma, lo almacena de manera adecuada y lo capitaliza para mejorar el objetivo del juego que es ganar.

Por otra parte, está la pregunta sobre la interacción y retroalimentación humana. Una máquina aprende de intercambios con humanos (u otros agentes) a través de retroalimentación. En el ajedrez puede ser la evaluación de una posición dada. Recordemos que dicha evaluación se obtiene con un conjunto de reglas, que al principio pueden ser una buena idea, pero que quizá después se muestre limitada. Por ejemplo, una buena regla podría ser capturar un peón sin ser atrapados, pues ganamos material (en términos de ajedrez, ganamos ventaja estática); ésta estrategia suele volver a las máquinas avariciosas al hacerlas capturar piezas cada pueden. | del rival para obtener un desarrollo más rápido de las piezas y atacarnos más rápido (en términos de ajedrez, el rival obtiene una ventaja dinámica). Esto es el sentido detrás de todas las aperturas de ajedrez denotadas como gambitos. En resumen, la idea es incorporar reglas a las máquinas, luego evaluarlas jugando contra personas (quienes evalúan una posición de manera distinta a las máquinas) e incorporar el conocimiento aprendido de dichas partidas y sus movimientos para que las máquinas tengan una mejor evaluación de las posiciones.

Descifrando emociones: clasificación de la soledad en diabéticos usando auto-machine learning

Pregunta: Gabriel Ordaz Baltazar.

Teniendo en cuenta que la depresión en diabéticos es existente debido a diferentes factores que son difíciles dentro del proceso de la enfermedad, ¿Es posible encontrar o existe una sección en el estudio que identifique en qué fase desde la adquisición o en su defecto del diagnóstico de la enfermedad es más probable que la persona comience a mostrar signos de depresión?

El enfoque principal de nuestro estudio fue investigar la clasificación de los niveles de depresión en pacientes diabéticos utilizando algoritmos de machine learning. Este trabajo se enfocó en la identificación de características cruciales que contribuyen a una clasificación de los niveles de depresión más precisa, que incluyen demográficas, resultados de laboratorio, historial de medicación, comorbilidades y respuestas a cuestionarios de salud mental. Sin embargo, el diseño del estudio no es ad hoc para un análisis que permita saber el momento exacto o la fase de la enfermedad en la que es más probable que esto ocurra. La recopilación y el análisis de nuestros datos se enfocaron en la identificación de características significativas en un momento dado, sin realizar un seguimiento longitudinal que permitiera discernir cambios a lo largo del tiempo desde el diagnóstico de la diabetes hasta la aparición de síntomas depresivos.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

Pese a ello, reconocemos la importancia de futuros estudios longitudinales, los cuales podrían revelar momentos importantes para la intervención temprana y la prevención de la depresión en esta población. Así, nuestros hallazgos actuales sobre las características asociadas a los niveles de depresión pueden ser una buena base para dichas investigaciones futuras.

Reconocimiento de brócoli mediante la anotación automática en imágenes RGB-D

Pregunta: Víctor Daniel Archundia Sánchez.

Con los resultados obtenidos y con la metodología que se aplicó para la detección de brócoli, ¿Cuáles son los principales beneficios y limitaciones de utilizar técnicas de aprendizaje automático en la agricultura, específicamente en la detección de cultivos como el brócoli?, y con esto mismo ¿Qué ventajas ofrece el uso de imágenes RGB-D en la detección y recolección automática de frutos o cultivos en comparación con otros métodos?

Primero responderemos la primera pregunta sobre los principales beneficios y limitaciones de utilizar técnicas de aprendizaje automático en la detección de cultivos como el brócoli. La cosecha automatizada en general presenta varios desafíos de investigación para la tecnología actual de percepción en tareas como la adquisición, preproceso, aprendizaje, detección y rastreo de los productos agrícolas a cosechar. Entre los desafíos se incluyen entornos complejos y muy poco estructurados, condiciones de iluminación difíciles, tasas variables de crecimiento de las plantas, así como incidencias climáticas (p. ej., temperatura, humedad, limitaciones de visibilidad por niebla o polvo, entre otros). Los entornos complejos requieren del desarrollo de dispositivos avanzados y robustos para cosechar selectivamente un cultivo de alto valor como el brócoli. Por lo tanto, desarrollar métodos automatizados para la cosecha selectiva capaces de identificar y separar con precisión la hortaliza ayudaría a aumentar la productividad y controlar mejor los costos de producción. Por ejemplo, un cultivo de brócoli debe ser cosechado con cuidado para no dañarlo, y en una etapa de madurez adecuada para mantener su calidad de nutrientes y frescura. Mientras que por un lado tubérculos y otros cultivos se han cosechado mecánicamente durante muchos años (principalmente porque pueden soportar un manejo menos delicado y aun así mantener su calidad), por otro los cultivos como el brócoli presentan desafíos diferentes debido a la naturaleza delicada del producto. Hoy en día la cosecha y el embalaje del brócoli todavía se realizan a mano en todo el mundo, con el consiguiente uso intensivo de mano de obra y de un proceso que requiere mayor tiempo, representando retos a vencer por tecnologías de machine learning en desarrollo.

En respuesta a la segunda pregunta sobre las ventajas que ofrece el uso de imágenes RGB-D en la detección y recolección automática de frutos o cultivos en comparación con otros métodos, podemos comentar lo siguiente. La detección 3D permitiría a una máquina cosechadora alcanzar niveles de autonomía con menos limitaciones, en contraste al uso exclusivo de visión 2D. Es decir, un dispositivo autónomo de cosecha de brócoli, equipado con sensores 3D, puede navegar en los campos cultivados y tomar decisiones de detección y corte en entornos complejos donde a los sensores 2D ya no les es posible. Un campo de cultivo presenta condiciones de iluminación variables y puntos de vista limitados que representan desafíos para los sensores 2D, pero que son manejables por sensores 3D. Aunado a lo anterior, están las ventajas del uso de información 3D, como el codificar la geometría de cada brócoli o el proporcionar diferentes puntos de vista en las condiciones poco estructuradas de los campos agrícolas. De manera análoga, utilizar únicamente técnicas de visión artificial 2D puede resultar insuficiente para un procesamiento fiable de las múltiples oclusiones que se producen a distintas profundidades en los cultivos de plantas de brócoli. Cabe mencionar en este punto que los sensores 3D actuales son asequibles, fiables y capaces de ofrecer datos RGB-D de alta resolución. Por otro lado, las necesidades computacionales y los desafíos algorítmicos pueden aumentar cuando se maneja información 3D, junto con las grandes cantidades de datos necesarios para entrenar detectores robustos; aunque esto también es prevalente con todos los dispositivos de adquisición actuales.



© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

Estado del IArte

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

IA++. Hacia un compromiso de uso razonable

En 2024 hemos sido testigos de la eclosión del uso de algoritmos de Inteligencia Artificial (IA) para realizar todo tipo de actividades, desde aquellas de la vida cotidiana hasta la exploración espacial. Los usuarios comunes tienen acceso, incluso sin solicitarlo, a las capacidades que ofrece la IA en aplicaciones tales como Whatsapp, Facebook, Instagram, Office por mencionar sólo algunos ejemplos. El uso que hacemos de dichas aplicaciones, permite que sigan aprendiendo y mejoran sus algoritmos. Lo que la mayoría de estos usuarios desconoce es que estas actividades - resumir un documento, generar una presentación o crear un sticker - tienen un costo energético y por ende un impacto al medio ambiente. El problema no es el acceso o la facilidad del uso de este tipo de aplicaciones para el público en general, ni las aplicaciones *per se*; sino la falta de compromiso entre los consumidores, los productores de tecnología IA y las reglas que rigen a ambos. En columnas anteriores hemos abordado el impacto al medio ambiente y costo energético que ocasiona el uso indiscriminado de las tecnologías[1], así como la necesidad de establecer criterios éticos y regular el desarrollo de tecnologías derivadas de algoritmos de IA [2].

El uso razonable de aplicaciones de IA depende de un compromiso entre tres actores: el usuario, las empresas y el gobierno. Como usuarios de aplicaciones debemos educarnos en el uso de la tecnología por convicción y no por facilidad, debemos evitar el abuso que se hace de estas tecnologías para realizar tareas humanamente sencillas. En lo que respecta a las empresas tanto desarrolladoras como consumidoras, Marvin Minsky[3] en el premio BBVA 2013 destacaba que se pueden crear máquinas tan inteligentes como las personas, teniendo los recursos económicos y el capital humano dedicado a la IA. Once años después de esta declaración, somos testigos de la lucha masiva entre las empresas que implementan los últimos avances en IA generativa. La revista Forbes[4] lanzó su sexto IA50 que reconoce a las empresas privadas de IA prometedoras, entre las que encontramos *OpenAI*, *Anthropic*, *Databricks* entre otras. También se destaca *Mistral*, considerado la competencia de *OpenIA*, un startup que desarrolla modelos de IA de código abierto y en su mayoría gratuitos que requieren menos procesamiento computacional. Mistral lanzó "*LeChat*" una alternativa que requiere menos procesamiento que *ChatGPT*. Entre las empresas que incorporan IA a sus aplicaciones existentes tenemos a *Google Workspace* y *Microsoft Office* como ejemplos. Por otro lado, las empresas consumidoras de tecnología IA deben mantener un equilibrio, entre los ingresos económicos y discernir entre utilizar la IA para potencializar a su capital humano o tomar la decisión de sustituirlos. Finalmente, el pasado 21 de marzo la Asamblea General de la ONU adoptó una resolución para regular los desarrollos tecnológicos en el área de IA. En particular, pide a los estados miembros : *que se abstengan de utilizar sistemas de inteligencia artificial que no pueden funcionar de conformidad con las normas internacionales de derechos humanos o los pongan en riesgo*[5].

Como personas pertenecientes a una sociedad, desde esta columna les invitamos a reflexionar y educarnos en el consumo razonable de aplicaciones con IA. Discriminar las tareas de nuestra vida cotidiana en las cuales es importante o crucial utilizar la IA; ser conscientes que toda consulta conlleva un costo energético y por ende un impacto al medio ambiente. El avance de la IA la podemos comparar con el sol. Siempre estará ahí, y así como hemos podido aprovecharlo, también hemos aprendido a protegernos, por lo que podríamos actuar de la misma forma con el uso de la IA.✷

[1] Hacia una IA sustentable - IArte - Año 14 (2022) número 3
[2] Ética para la IA - IArte - Año 15 (2023) número 3
[3] https://shorturl.at/51kZL
[4] https://www.forbes.com.mx/lista-forbes-ia-50-estas-son-las-principales-empresas-emergentes-de-inteligencia-artificial/
[5] https://news.un.org/es/story/2024/03/1528511

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

Sakbe

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

IA y los desafíos de la Humanidad

La ONU y el cambio climático ⇒ La ONU está utilizando la IA para abordar el cambio climático a nivel mundial, y un ejemplo es la aplicación MyAnga en Kenia. Desarrollada por Amfratech, este software emplea IA para proporcionar datos meteorológicos y satelitales a los pastores a través de sus dispositivos móviles, permitiéndoles prepararse para sequías y gestionar mejor su ganado. Disponible en Google Play de forma gratuita en inglés y swahili, MyAnga ofrece observaciones meteorológicas diarias, pronósticos para los próximos siete días, distribución de precipitaciones durante los últimos 30 días, y condiciones de humedad y sequía esperadas. Con un diseño intuitivo, se convierte en una herramienta valiosa para mejorar la resiliencia ante los desafíos climáticos en curso.

https://news.un.org/en/story/2023/11/1143187

https://ustadi.org/2020/07/09/myanga-app-a-game-changer-for-pastoralist-climate-resilience/

World Bank y la gobernanza ⇒ El Banco Mundial emplea IA para mejorar la gobernanza. En Bangladesh, un estudio reciente utilizó el aprendizaje automático (ML, por su vocablo inglés Machine Learning) para identificar inversiones estratégicas que aceleren el acceso a servicios de salud materna, neonatal, infantil y de nutrición (RMNCH-N). Este enfoque destaca la importancia de la disponibilidad de personal capacitado y la calidad de la atención para aumentar la utilización de servicios, con intervenciones que podrían prevenir hasta el 95 % de las muertes por diarrea y el 67 % de las muertes por neumonía en niños menores de 5 años a nivel global para el año 2025. Sin embargo, aún se necesita más investigación para comprender completamente las relaciones causales entre los determinantes de la demanda y los resultados de RMNCH-N.

https://blogs.worldbank.org/en/developmenttalk/using-machine-learning-unravel-health-utilization-drivers-bangladesh

IEA y la gestión energética ⇒ IEA y la gestión energética La IEA trabaja con gobiernos y la industria para moldear un futuro energético seguro y sostenible para todos. Un caso de estudio reciente sobre el uso de IA en sistemas de gestión energética de edificios comerciales destaca la forma en que esta tecnología puede optimizar el consumo de energía en tiempo real. Por ejemplo, el sistema Flex2X desarrollado por Grid Edge utiliza algoritmos de IA para analizar datos de consumo energético, tomar decisiones en tiempo real y participar en mercados de demanda flexible. Aunque esta tecnología promete importantes beneficios económicos y ambientales, aún existen barreras regulatorias y de confianza que deben superarse para su adopción generalizada.

https://www.iea.org/articles/case-study-artificial-intelligence-for-building-energy-management-systems

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

ARTÍCULO ACEPTADO

¿A dónde vas?: Aplicaciones de la IA a la Predicción del Movimiento de Peatones

Jean Bernard Hayet, David Gómez Gutiérrez, Leobardo Campos Macías y Javier Felip León

Anticipación para interacciones seguras

Los avances de la Inteligencia Artificial (IA) en la última década han sido espectaculares y sorprendentes, incluso para la mayoría de los profesionistas del área. En este artículo, queremos dar cuenta de cómo algunas de las técnicas de aprendizaje de máquina desarrolladas en estos años han permitido enormes progresos en el área de la *predicción del movimiento de agentes humanos así como de sus intenciones*. Nos enfocaremos en particular en el problema de predicción que surge en la navegación autónoma de coches, una de las aplicaciones de mucha actualidad, donde esta tarea de inferir cuáles pueden ser los movimientos futuros y las acciones de un humano observado por una máquina es de vital importancia.

El desarrollo de coches autónomos tiene el potencial de disminuir drásticamente el número de accidentes automotrices, ya que la gran mayoría de estos son el producto de errores y descuidos humanos. A pesar de ser un problema tecnológico muy complejo, donde una falla puede tener consecuencias catastróficas, coches autónomos se encuentran a prueba en distintas partes del mundo.

Uno de los mayores retos para la adopción amplia de esta tecnología es garantizar la coexistencia *segura* con usuarios "frágiles", como peatones y bicicletas. Para lograrlo, no basta detectar y clasificar objetos en movimiento entre los datos perceptuales, sino que se requiere también determinar las intenciones de los usuarios y predecir sus (probables) siguientes movimientos para *anticipar* la evolución de la escena y preparar acciones adecuadas (p. ej., parar, ralentizar, dar la vuelta...).

Esta tarea de predicción se hace a partir del entendimiento de la escena y de una observación del comportamiento del usuario en un corto lapso de tiempo (típicamente, algunos segundos), a través de datos sensoriales. En muchas ciudades, esos coches tienen un alto nivel de interacción con peatones y bicicletas, en ambientes no estructurados. Muchos de estos casos son retadores, ya que requieren tomar decisiones alrededor de usuarios cuyo comportamiento es difícil de predecir.

En alguno de estos escenarios, para lidiar con el reto de una predicción basada en una observación corta y con entendimiento del contexto, se pueden utilizar cámaras en la infraestructura (por ejemplo, en una intersección de tráfico inteligente que comunica las trayectorias y/o las predicciones a los vehículos que se aproximan) para que con esta información se tomen decisiones.

Este artículo describe diferentes esfuerzos para producir sistemas eficientes de predicción de movimientos de peatones aprovechando los avances de la IA, sus alcances y limitaciones.

Planteamiento del problema

A manera de ilustración, en la Figura 1, describimos el planteamiento del problema de predicción de movimiento como se trata generalmente en la literatura. Se consideran escenas monitoreadas por sensores, típicamente cámaras, que pueden estar ubicados a bordo de un vehículo (por ejemplo en el caso de coches, como en la Figura 2) o estáticos en la escena (por ejemplo, cámaras de vigilancia o desde drones, como en la Figura 3).

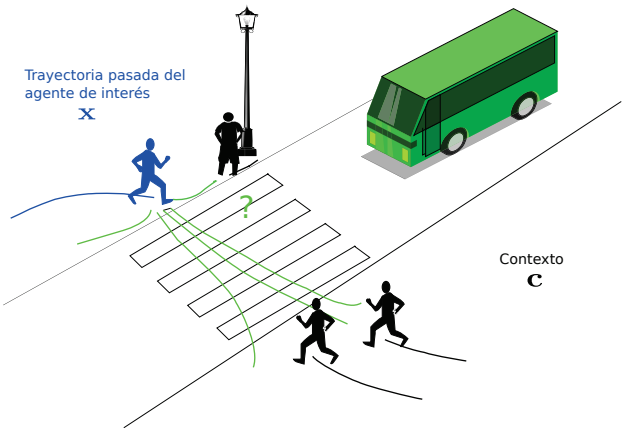


Figura 1. El problema de predicción de trayectoria de un agente de interés (en azul) con contexto **c**.



Figura 2. Ejemplo de observaciones tomadas desde un vehículo (imagen del dataset JAAD[7]).

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

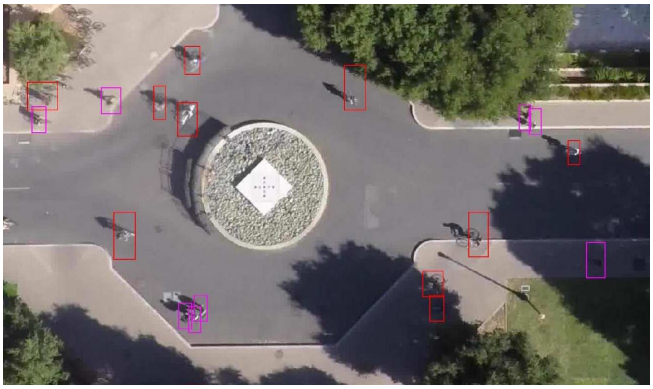


Figura 3. Ejemplo de observaciones tomadas desde un dron estático (imagen del dataset SDD[8]).

Dentro de lo observado por los sensores, nos enfocamos en las entidades *potencialmente en movimiento*, por ejemplo peatones, coches y ciclistas. Los algoritmos de detección de objetos en imágenes[7] y de rastreo visual son ahora muy eficaces y permiten generar observaciones sobre estas entidades (los cuadros visibles en las Figuras 2 y 3), es decir, rastrearlas hasta cierto momento y estimar su trayectoria. La tarea de predicción de movimiento es la siguiente: Dadas *observaciones* realizadas durante cierto tiempo por esos sensores y algoritmos de rastreo, y dado un agente de interés en especial (por ejemplo, el personaje azul de la Figura 1) que tiene una trayectoria estimada que llamamos $\mathbf{x}$ (la curva azul precediendo el personaje), y dado un *contexto* que nos da más información sobre la situación en que ocurre el movimiento y que llamamos $\mathbf{c}$, queremos estimar *cómo sigue el movimiento del agente después de* $\mathbf{x}$, es decir, cuál es su trayectoria futura que llamamos $\mathbf{y}$.

En otras palabras, se trata de estimar qué trayectoria futura explicaría mejor las observaciones actuales; algunos ejemplos de posibles trayectorias futuras son las que aparecen en verde.

El contexto $\mathbf{c}$ puede ser de diferente naturaleza según la aplicación: puede incluir información sobre el movimiento de todos los otros agentes de la escena; puede indicarnos, a través de un *mapa* cuáles son los elementos semánticos que rodean a nuestro agente, por ejemplo la estructura en la cual evolucionan los agentes de la Figura 3 (con el parque en verde, la rotonda, la carretera).

En primer acercamiento, predecir el movimiento de agentes se puede plantear matemáticamente como determinar una *función*, que llamaremos π, que a un pasado $\mathbf{x}$ y un contexto $\mathbf{c}$ asocia una trayectoria futura $\mathbf{y}$:

$$\mathbf{y} = \pi(\mathbf{x}, \mathbf{c}).$$

Este tipo de problema se conoce como de *regresión*.

Un reto mayor para la adopción amplia de coches autónomos, es garantizar la coexistencia segura con usuarios "frágiles", como peatones y bicicletas. Para eso, la predicción del movimiento es fundamental.

El papel del enfoque probabilista

En la ilustración de la Figura 1, vemos que varias cosas podrían ocurrir: que el peatón vaya a conversar con la persona parada bajo la luminaria; que dé la vuelta; que cruce la calle en el pasaje peatonal, evitando al par de personas que se aproximan en sentido contrario. Cada escenario corresponde a una posible trayectoria en verde.

Una limitante de plantear el problema de predicción de trayectorias como "simple" problema de regresión, como descrito arriba, es que recuperamos de salida de nuestro modelo un *solo* valor $\hat{\mathbf{y}}$ para la predicción, entonces, por definición, solo producimos una hipótesis sobre el futuro de la trayectoria. Ahora, en este problema, a menudo la información que recibimos de entrada (trayectoria pasada y contexto) deja *incertidumbre* en lo que puede ocurrir después, como en la situación de la Figura 1. En este caso, esperaríamos que nuestra predicción sea *multimodal*, es decir, que refleje la incertidumbre subyacente a la situación y que sea capaz de producir múltiples hipótesis. De forma más general, en la predicción de intenciones de humanos participan tres fuentes principales de incertidumbre: el error de las medidas de los sensores, el error entre el proceso modelado y la realidad y la aleatoriedad e impredecibilidad del comportamiento humano. Cada ser humano es diferente y su toma de decisiones es incierta a ojos de un sistema de predicción.

Por ende, es necesario considerar múltiples hipótesis de comportamiento posibles y asignar una probabilidad a cada una de ellas, basados, por ejemplo, en experiencias previas. De esa manera, podemos crear sistemas de IA que pueden ser confiables y proporcionan resultados interpretables. Además, el enfoque probabilista abre la puerta al uso de técnicas Bayesianas, las cuales consideran un modelo generativo (generador de hipótesis), el cálculo de la verosimilitud (es decir, estimar numéricamente qué tanto los datos observados están congruentes con una hipótesis), y cómo generar la siguiente hipótesis basada en la anterior.

Predicción usando modelos físicos

Una manera simple de predecir la trayectoria de un agente es utilizar la trayectoria pasada $\mathbf{x}$ y hacer una supo-

sición física simple sobre la trayectoria: Por ejemplo, suponer que el agente mantendrá una velocidad constante, o una aceleración constante. Este modelo físico sencillo resulta útil en casos simples, sin embargo, hay dos aspectos clave que no considera: el contexto y la incertidumbre (ya que se tratan de modelos *determinísticos*). Para considerar el contexto, se debe incorporar el cruce peatonal, estado de los semáforos, densidad del tráfico, presencia de vallas u obstáculos y otros elementos que tienen una influencia directa en la toma de decisiones. Existen trabajos que utilizan simuladores de tráfico como modelos para considerar los aspectos del contexto[2].

Otro aspecto importante a tener en cuenta son las dinámicas de los peatones y cómo interactúan entre sí. En este ámbito, existen trabajos que son capaces de modelar correctamente el efecto de las interacciones entre individuos y grupos de peatones utilizando variantes de un modelo conocido como fuerzas sociales[4]. Estos modelos toman un enfoque de física clásica, al modelar cada agente como una partícula sobre la cual se ejercen fuerzas de repulsión o de atracción por parte de los otros agentes o por parte de elementos contextuales; se pueden refinar los parámetros de estos modelos por cada agente para ajustarse a sus más recientes observaciones. Ahora, si bien pueden explicar ciertas trayectorias y simular lo que viene, siguen siendo modelos determinísticos.

Predicción como planificación inversa

La planificación de movimientos tiene como objetivo generar una trayectoria desde una ubicación hacia una o varias metas, en un ambiente conocido o parcialmente conocido. Esta técnica es normalmente utilizada en la navegación de robots para obtener una secuencia de movimientos para alcanzar la meta. El problema de predicción de movimiento puede ser considerado como una planificación inversa[1], ya que, a partir del conocimiento de la escena y de una porción de la trayectoria, se busca la meta que mejor explica las observaciones. En este sentido, para predecir intenciones, se pueden utilizar planificadores de movimiento para generar hipótesis hacia las posibles metas en la escena y asignar probabilidades basadas en que tan congruentes son las trayectorias generadas con respecto a las observaciones.

La Figura 4 muestra un ejemplo de la aplicación de la planificación inversa para inferir las posibles trayectorias y metas de un agente de interés. Primero, se considera la información previa del ambiente, en este caso, los obstáculos (edificios) y posibles lugares donde los agentes de interés transitan con menor o mayor frecuencia, por ejemplo, los pasos peatonales. Después, dada la observación parcial de una trayectoria de un agente de interés, se infieren posibles caminos a diferentes metas utilizando métodos de planificación. Para el ejemplo de la Figura 4, se utilizó el algoritmo de planificación llamado Probabilistic Roadmap (PRM)[5].

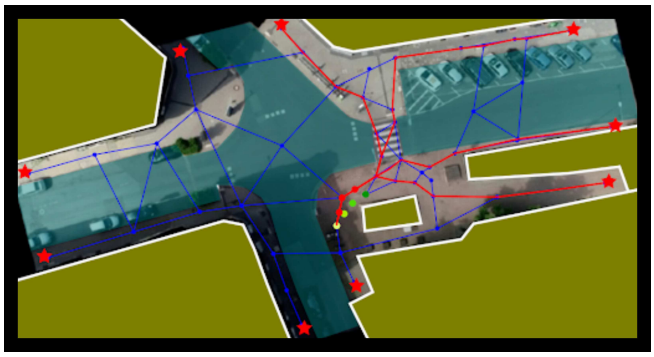


Figura 4. Ejemplo de planificación inversa; muestra, para diferentes metas (estrellas rojas), las rutas (líneas rojas) congruentes y la trayectoria observada del peatón (puntos en amarillo-verde en el centro de la imagen, degradados según el tiempo de observación).

Este tipo de esquemas permite tomar en cuenta el contexto de una forma natural al considerar, por ejemplo, el estado del semáforo, los vehículos que se aproximan, las obstrucciones y las interacciones entre usuario. Esto se realiza sin la necesidad de tener algún entrenamiento previo, en contraste con los métodos basados en aprendizaje. Sin embargo, al considerar más condiciones en el ambiente, el problema de planificación se torna más complejo de resolver. Además, las trayectorias generadas por estos algoritmos tienden a ser diferentes a las que adopta un ser humano.

Modelos basados en datos

Desde el final de la década de los 2010s y con el renacimiento de la IA con enfoque conexionista, la predicción de movimiento ha tomado un giro fuerte hacia el uso de modelos *basados en datos*. Bajo este esquema, la predicción de trayectorias se plantea como un problema de aprendizaje automático, más específicamente, de *regresión*. Significa que se intenta determinar una función parametrizada que a una entrada (en este caso, los datos de una trayectoria pasada $\mathbf{x}$ y los del contexto asociado $\mathbf{c}$) asocia una salida $\hat{\mathbf{y}}$. Para eso, dada una base de datos de ejemplos (típicamente, miles) de trayectorias pasadas $\mathbf{x}$, contextos asociados $\mathbf{c}$, y trayectorias "futuras" $\mathbf{y}$, ajustamos los parámetro θ de nuestra función parametrizada π_θ (también llamado "modelo predictivo") de tal manera que la trayectoria futura predicha $\hat{\mathbf{y}}$ se acerque a la que está presente en nuestros datos de entrenamiento. Un criterio de optimización de esos parámetros puede ser de la forma siguiente:

$$\theta^* = \arg\min_\theta \sum_{\mathbf{x},\mathbf{c},\mathbf{y}} \|\mathbf{y} - \pi_\theta(\mathbf{x},\mathbf{c})\|^2.$$

Una vez que el modelo está entrenado (es decir, que hemos encontrado parámetros óptimos θ^*), si recibimos nuevos datos de trayectorias pasadas y contextos $\mathbf{x},\mathbf{c}$

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

entonces podemos predecir la continuación de la trayectoria al aplicar $f_{\theta^*}(\mathbf{x}, \mathbf{c})$. Como podrán imaginar, en muchos casos f_θ está implementada como una red neuronal, y, dada la naturaleza *secuencial* de los datos ($\mathbf{x}$ es una secuencia espacio-temporal de puntos), las redes más utilizadas son las redes recurrentes, como las *Long Short-Term Memory* (LSTM) o las *Gated Recurrent Unit* (GRU) y variantes de redes transformadoras. Tienen la ventaja de generar, por construcción, trayectorias naturales (parecidas a los datos de entrenamiento); ahora, no permiten per se el manejo de las incertidumbres.

El papel clave de la IA generativa

Hemos mencionado arriba el poder realizar una predicción que refleje la incertidumbre del problema y capaz de producir múltiples hipótesis. Este tipo de contexto es en el que sobresalen los *modelos generativos*. Son conocidos por sus aplicaciones populares a la generación de imágenes o vídeos sintéticos realistas a partir de texto (DallE o *Stable Diffusion*, por ejemplo).

Esencialmente, los modelos generativos actúan como *sustitutos a distribuciones de probabilidad*: se entrenan con ejemplos de muestras de una cierta distribución de probabilidad y, una vez entrenados, serán capaces de generar *nuevas* muestras que, estadísticamente, comparten las caractéristicas de los ejemplos vistos durante el entrenamiento. Entre los modelos generativos más populares podemos citar a las redes generativas adversariales (GANs), los auto-encodificadores variacionales (VAEs)[9] y, más recientemente, los modelos de difusión.



Figura 5. Ejemplo de predicción realizada por un modelo generativo, el VAE (imagen de los autores).

En el caso de predicción de movimiento, una idea presente en muchos trabajos recientes es usar estos modelos generativos de forma condicional, es decir, para sustituir a la *distribución de probabilidad condicional* $p(\mathbf{y}|\mathbf{x}, \mathbf{c})$ y producir muestras $\mathbf{y}^{(k)}$ de ella, cada uno pudiéndose interpretar como una hipótesis de movimiento. En la Figura 5, damos un ejemplo de predicción realizada por un modelo generativo (un VAE, similar al descrito en[9]): las líneas cian representan centenas de hipótesis que hace el modelo sobre la trayectoria futura del peatón con trayectoria observada azul $\mathbf{x}$ y contexto $\mathbf{c}$, el conjunto de trayectorias de los otros agentes. La extensión espacial y la densidad de estas hipótesis nos dan una cuantificación de la incertidumbre en la predicción.

Evaluación de los modelos predictivos

Una cuestión importante al momento de desarrollar un modelo predictivo como los que hemos descrito arriba es cómo evaluamos si un modelo es mejor que otro. Las métricas más comunes se llaman Error de Desplazamiento Promedio (ADE) y Error de Desplazamiento Final (FDE). Ambas calculan la distancia Euclidiana entre las posiciones predichas y verdaderas, y la primera hace el promedio de estos errores dentro de la ventana temporal de predicción, mientras la segunda calcula el error en el último paso de esa misma ventana. En el caso de modelos de múltiples hipótesis (como los que usan modelos generativos), el estándar es generar H hipótesis y evaluar los valores mínimos del ADE y FDE entre estas H trayectorias predichas. Con un sistema del estado del arte que desarrollamos recientemente[9], con $H = 20$, y para el tipo de situación de la Figura 5, con horizonte de tiempo de $4.8s$, los errores FDE, por ejemplo, varían entre $20cm$ y $80cm$.

Predecir intenciones en otros ámbitos

El tipo de enfoques que hemos mencionado hasta ahora tienen también otras aplicaciones. Actualmente las líneas de manufactura ya incorporan robots como parte del proceso de producción. Sin embargo, las tareas que realizan humanos y robots suelen estar separadas. Con el reciente avance de la IA, los robots se pueden encargar de más tareas repetitivas y que no requieren de cognición avanzada. Por esto, hay una necesidad de desarrollar sistemas robóticos capaces de colaborar con humanos en tareas de manipulación y ensamblado.

Los métodos de predicción de trayectorias de peatones, poseen competencias transversales que pueden ser aplicadas a otros dominios de la robótica, como por ejemplo la colaboración humano-robot. En este ámbito, en [3] se presenta un sistema que, gracias a la predicción de intenciones, es capaz de mejorar múltiples aspectos de la colaboración entre humanos y robots cuando se realizan tareas de manipulación en espacios compartidos.

En la Figura 6, se muestra como un humano y un robot pueden colaborar con fluidez. Esto se consigue gracias a que el sistema robótico monitorea y predice los movimientos del humano. La tarea, consistente en empacar todos los objetos de la mesa en la caja, puede resolver-

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

se más eficientemente anticipando qué objeto el humano quiere alcanzar. La predicción de intenciones (visualizada en la pantalla del segundo plano) permite al robot anticipar qué objeto el humano está tratando de manipular para agarrar otro que esté alejado. Maximizando el paralelismo de las acciones y evitando interferencias entre ambos agentes. De esta forma, el robot no necesita esperar a que el humano se retire y puede simultáneamente agarrar y empacar otro de los objetos que están alejados. Este tipo de colaboración fluida y con comunicación meramente no-verbal es propia de los humanos. El desarrollo de tecnologías que permitan a los sistemas robóticos interactuar con humanos de forma más natural abre la puerta a un mundo de tareas híbridas donde humanos y robots colaboren aportando cada uno sus características diferenciadoras. Por ejemplo, en caso del robot, fiabilidad, precisión, repetibilidad y resistencia. En caso del humano, flexibilidad, raciocinio, maniobrabilidad.

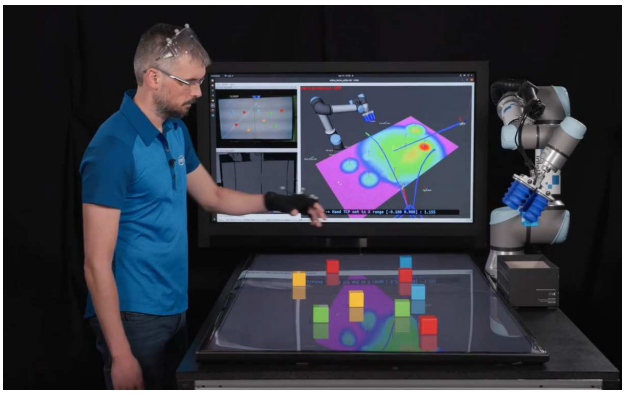


Figura 6. Humano y robot colaboran sobre una superficie compartida.

La predicción de trayectorias se puede plantear como un problema de regresión permitiendo asociar a una trayectoria pasada y su contexto una trayectoria futura.

Tendencias actuales y futuras

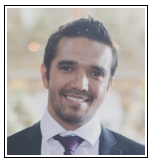
Si bien los avances de los últimos diez años han sido espectaculares, el uso de la IA para la predicción de movimiento humano tiene todavía varios retos que resolver. Primero, la mayoría de los sistemas basados en datos sufren de la falta de interpretabilidad en sus salidas; una tendencia muy reciente para atacar este problema es usar sistemas híbridos que combinan la eficiencia de los sistemas basados en aprendizaje profundo con los modelos matemáticos que aportan mayor interpretabilidad[10]. Otra área de mejora es la realización de sistemas más robustos. En la actualidad, la casi totalidad de los sistemas propuestos usan bases de datos limpias y sin errores de rastreo, lo que no corresponde a condiciones reales; además, no están todavía adaptados a altas densidades de agentes. Por ende, podemos apostar a que la próxima década nos traerá más desarrollos hacia la anticipación de las acciones humanas por parte de máquinas.✷

REFERENCIAS

1. Baker, C. L., Saxe, R., y Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, 113(3):329–349.
2. Felip, J., Ahuja, N., G´omez-Gutiérrez, D., Tickoo, O., y Mansinghka, V. (2019). Real-time Approximate Bayesian Computation for Scene Understanding.1–15. arXiv:1905.13307.
3. Felip, J., Gonzalez-Aguirre, D., y Nachman, L. (2022). Intuitive and efficient human-robot collaboration via real-time approximate bayesian inference. En *2022 IEEE/RSJ International Conf. on Int. Robots and Systems* (IROS), 3093–3099.
4. Helbing, D. y Moln´ar, P. (1995). Social force model for pedestrian dynamics. *Physical review. E, Statistical physics, plasmas, fluids, and related interdisciplinary topics*, 51:4282–4286.
5. Kavraki, L. E., Svestka, P., Latombe, J.-C., and Overmars, M. H. (1996). Probabilistic roadmaps for path planning in high-dimensional configuration spaces. *IEEE transactions on Robotics and Automation*, 12(4):566–580.
6. Rasouli, A., Kotseruba, I., y Tsotsos, J. K. (2017). Agreeing to cross: How drivers and pedestrians communicate. En *IEEE Intelligent Vehicles Symposium* (IV), 264–269.
7. Redmon, J. y Farhadi, A. (2017). Yolo9000: Better, faster, stronger. En *2017 IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 6517–6525.
8. Robicquet, A., Sadeghian, A., Alahi, A., y Savarese, S. (2016). Learning social etiquette: Human trajectory understanding in crowded scenes. En Leibe, B., Matas, J., Sebe, N., and Welling, M., editors, Computer Vision – ECCV 2016, 549–565.
9. Xu, P., Hayet, J.-B., y Karamouzas, I. (2022). Socialvae: Human trajectory prediction using timewise latents. En *European Conf. on Computer Vision*, 511–528. Springer.
10. Yue, J., Manocha, D., y Wang, H. (2022). Human trajectory prediction via neural social physics. En *The European Conference on Computer Vision* (ECCV), 376–394.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

SOBRE LOS AUTORES



David Gómez Gutiérrez obtuvo el grado de doctor en Ciencias en Ingeniería Eléctrica en Cinvestav, Unidad Guadalajara en 2013. Es investigador en el Intelligent Systems Research Lab, Intel Labs, miembro del Tecnológico Nacional de México, Instituto Tecnológico José Mario Molina Pasquel y Henríquez y miembro SNII Nivel II. Sus intereses científicos incluyen la navegación autónoma de multi-robots, los algoritmos de IA para la predicción de movimiento y el control robusto.



Jean-Bernard Hayet obtuvo un grado de ingeniería en la Ecole Nationale Supérieure de Techniques Avancées (ENSTA) en París en 1999, su grado de maestría en IA por la Universidad París VI y su grado de doctorado por la Universidad de Toulouse en 2003, con una tesis en robótica autónoma. Trabajó como post-doc en la Universidad de Liejas (Bélgica) desde 2004 a 2007 en temas de visión artificial. Desde 2007, es investigador en el CIMAT, en Guanajuato, México. Realiza trabajos de investigación en aplicar técnicas de aprendizaje automático en visión artificial y robótica autónoma.



Leobardo Campos Macías es científico investigador en robótica en Intel Labs. Ha desarrollado e implementado algoritmos para la navegación autónoma en entornos desconocidos. Actualmente, desarrolla métodos para la colaboración segura entre humanos y robots. Recibió su maestría y doctorado por el Cinvestav investigando robots humanoides y autonomía de drones, respectivamente. Es miembro del Sistema Nacional de Investigación, tiene más de 50 patentes en trámite y fue ganador del premio a la mejor tesis de doctorado en el 2022 otorgado por la Academia Mexicana de Ciencias.



Javier Felip León es científico investigador en Intel Labs desde 2016. Obtuvo su doctorado en Ciencias de la Computación de la Universitat Jaume I (España) en 2016. Su trabajo de doctorado se centró en la manipulación robótica bajo incertidumbre, aportando a este campo con perspectivas prácticas. Actualmente, su investigación se centra en métodos de análisis por síntesis, métodos probabilísticos aplicados y colaboración humano-robot.

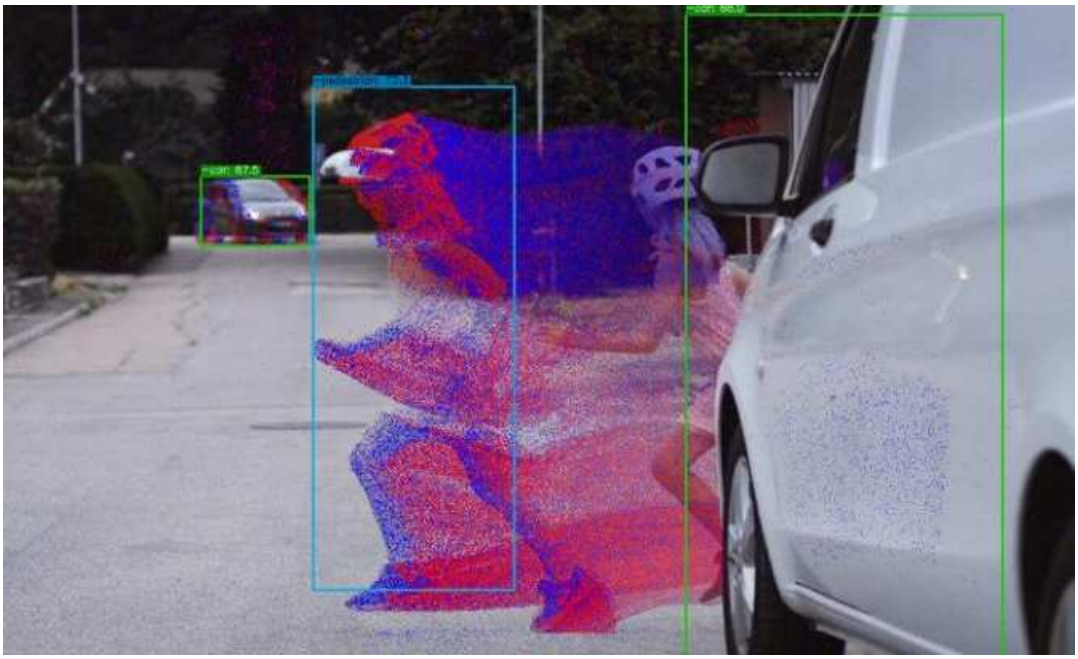


Imagen tomada de:
https://noticiasdelaciencia.com/art/50822/nuevo-sistema-de-inteligencia-artificial-para-detectar-peatones-y-evitar-atropellos.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

ARTÍCULO ACEPTADO

Monitoreo inteligente para el cuidado de los adultos mayores en sus hogares

Elizabeth López Lozada, Humberto Sossa Azuela y Elsa Rubio Espino

Introducción

A nivel mundial, la población de adultos mayores crece a un ritmo acelerado. En 2022, la Organización Mundial de la Salud publicó que había 1,400 millones de adultos mayores y prevé que para 2050 la cifra incremente a 2,100 millones [1]. De acuerdo con las estadísticas del INEGI [2], la cantidad de personas mayores de 60 años, en México, asciende a 17.9 millones.

Adicional a esto, este organismo destaca que en los últimos años los adultos mayores de 60 años tienden a vivir de forma independiente con sus parejas o en hogares unipersonales (hogares en donde vive solo una persona). Esto es algo alarmante, ya que si la población de adultos mayores va en aumento, estas tendencias crecerán. Dicha situación implica algunos riesgos; por ejemplo, en caso de accidentes, problemas de salud o eventos anormales que requieren atención, los adultos podrían quedarse a su suerte si nadie notifica o atiende la situación. Por ello, se enfatiza la importancia del desarrollo de tecnologías que realicen el seguimiento y monitoreo automático de las personas para determinar anomalías en su comportamiento, detectar trastornos físicos o mentales e incluso situaciones de alto riesgo.

Hoy en día, existen diferentes productos en el mercado para asistir a los adultos mayores con tareas cotidianas, como recordatorios para la toma de medicinas, botones de ayuda, la medición del ritmo cardíaco, saturación de oxígeno en la sangre, glucosa y peso, localización por GPS y detección de caídas [3, 4]. Estos productos requieren que las personas utilicen dispositivos externos que se colocan en el cuerpo para realizar las mediciones o que siempre carguen con sus teléfonos móviles, cosa que podrían olvidar.

Las tecnologías de videovigilancia permiten realizar el monitoreo remoto de las personas utilizando cámaras de video fijas que se colocan en lugares estratégicos dentro del hogar. Esto sirve de apoyo a los adultos mayores a vivir independientemente y a sus familiares o cuidadores a realizar un monitoreo remoto de ellos, hacer el seguimiento de sus actividades y detectar si existen comportamientos anómalos, sin la necesidad de que lleven consigo dispositivos extras.

Actualmente, las tareas de videovigilancia se siguen haciendo de forma manual, con operadores o familiares que se dedican a observar lo que ocurre en los videos y responden de acuerdo a la situación. No está por demás decir que esta es una labor intensiva, que puede presentar fallas por alguna distracción o error humano. En este sentido, el desarrollo de tecnologías que automaticen esta tarea es todo un reto que puede solucionarse con ayuda de la Inteligencia Artificial (IA). Hay grandes expectativas de la IA en el ámbito de la salud, y ende para los cuidados de la salud, gracias al desarrollo de programas de computadora que permiten analizar datos y detectar patrones y tendencias que pueden indicar problemas de salud o comportamientos fuera de lo normal en las personas.

Los sistemas inteligentes de videovigilancia permitirán asistir a las personas cuando sufrán accidentes en casa.

Monitoreo de video

El monitoreo de video consiste en la vigilancia de sistemas compuestos por cámaras de video, dispositivos de almacenamiento y dispositivos para la transmisión de las imágenes. Se conocen como sistemas de videovigilancia, que permiten observar y registrar lo que ocurre en las calles, las oficinas, los centros comerciales y los hogares. Son una herramienta útil en el ámbito de la seguridad para la prevención y detección de delitos, aunque no es la única aplicación que tienen.

Los sistemas de videovigilancia se pueden dividir en dos categorías: videovigilancia tradicional y la videovigilancia o monitoreo inteligente. Para ambos casos, se coloca un conjunto de cámaras de video para monitorear lo que ocurre en un área específica.

La videovigilancia tradicional [5] consiste en que las cámaras registren la actividad que está ocurriendo dentro del área de cobertura y que una persona vigile las cámaras constantemente, como se muestra en la Figura 1. Este tipo de videovigilancia se realiza de forma manual, porque se necesita que alguien esté siempre observando el video para detectar los eventos de interés.

Por otro lado, en la videovigilancia inteligente, mostrada en la Figura 2, todo lo recogido por las cámaras

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

es procesado por un sistema computacional que registra las actividades o detecta los eventos deseados. Dicho sistema computacional lo forman programas y algoritmos de IA que permiten registrar el video, detectar actividades o eventos y automatizan el envío de notificaciones o encendido de alarmas.



Figura 1. Sistema de videovigilancia tradicional.

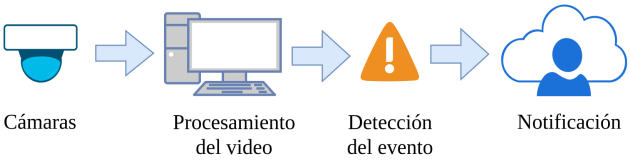


Figura 2. Sistema de videovigilancia inteligente.

Este tipo de monitoreo puede ayudar a mejorar la vida de sus habitantes cuando son capaces de detectar, anticipar y responder a sus actividades diarias, ayudándoles de forma apropiada y oportuna, puesto que es un tipo de tecnología que se puede adaptar bien para el análisis de video y el cuidado de los adultos mayores.

Monitoreo de inteligente de actividades

Si analizamos lo que implica el cuidado de las personas, podríamos decir que el reconocimiento y la detección de actividades es una tarea esencial para ello.

Dentro del hogar, la mayoría de las personas tienen una rutina que consiste en realizar ciertas actividades de forma cotidiana. Por ejemplo, la rutina puede consistir en levantarse, bañarse, desayunar y salir al trabajo o la escuela; actividades que ocurren de dos a más días por semana y que podemos reconocer como un comportamiento normal. Sin embargo, si alguien se queda en cama durante todo el día, es un comportamiento inusual que podría indicar de que la persona no se siente bien. Esto se puede detectar automáticamente utilizando técnicas de reconocimiento de actividades en video.

El reconocimiento de actividades humanas (RAH) consiste en que una computadora identifique, por medio de algoritmos de IA, las actividades que se observan en un video. De este modo, el sistema es capaz de determinar si una persona está mirando el televisor o está leyendo un libro. El esquema general, en la Figura 3, del funcionamiento del RAH está formado por tres bloques principales. Durante el preprocesamiento de datos se aplican métodos para mejorar la calidad las imágenes, recortar o cambiar el tamaño del video, resaltar detalles de las imágenes, etc.

El bloque del entrenamiento se compone por los algoritmos y los modelos de IA que extraen características y las clasifican; un ejemplo de estos modelos son las llamadas redes neuronales. El entrenamiento es una etapa que consiste en mostrarle al modelo muchos ejemplos de actividades y que aprenda a que un video corresponde a la actividad de "leer". A partir de los resultados del entrenamiento, un bloque del sistema se encargará de enviar la actividad reconocida. En el ejemplo de la Figura 3, el sistema de RAH reconoce al video de entrada con la actividad de "Mira la TV".

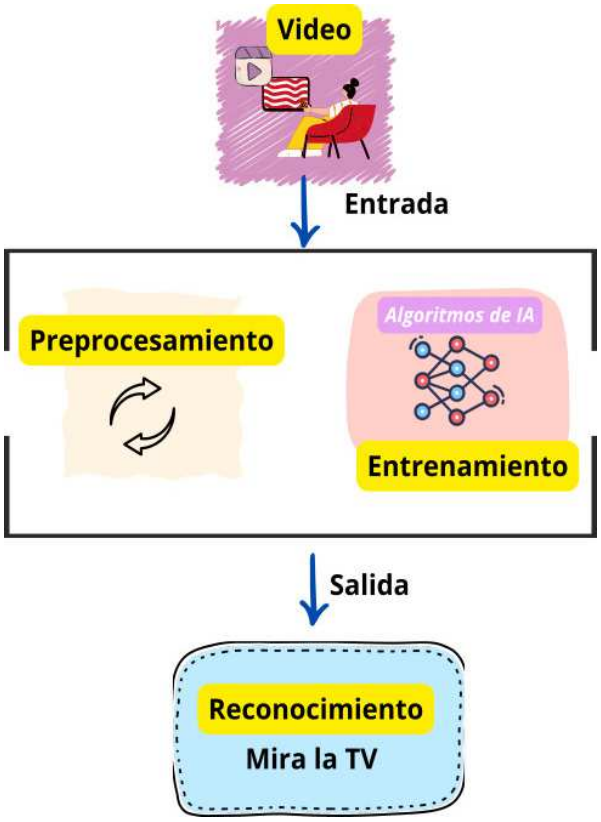


Figura 3. Diagrama general del RAH.

Dado el desarrollo de la IA, sobre todo hablando del RAH, se observa que integrar la videovigilancia con el reconocimiento de actividades y el análisis de video es de utilidad para determinar el comportamiento de las personas que se están monitoreando. Esto podría ayudar a automatizar el monitoreo remoto y permitir que los familiares o cuidadores de la salud reciban alertas sobre algún comportamiento fuera de lo normal. También ayudan a detectar eventos que requieren de una asistencia médica inmediata, como por ejemplo, que la persona caiga el suelo y no logre levantarse.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

Un modelo de reconocimiento de actividades humanas

Motivados por lo anterior, en el Laboratorio de Robótica y Mecatrónica del Centro de Investigación en Computación del Instituto Politécnico Nacional comenzamos a trabajar con una propuesta para el reconocimiento de acciones humanas en video utilizando algoritmos de IA, con el fin, de desarrollar un sistema de monitoreo inteligente que puede ser usado para el cuidado de los adultos mayores desde sus hogares. La propuesta inicial consiste en detectar y seguir a las personas en el video utilizando un seguidor de personas llamado FairMOT [6]. Una vez que la persona es detectada, las imágenes del video se redimensionan para ingresarlas al extractor de características. La extracción de las características del video se realiza con un modelo previamente entrenado, llamado ConvNeXt [7], con el cual se extraen las características principales de cada una de las imágenes que forman el video y se ingresan al modelo de clasificación.

Tabla 1. Resultados experimentales.

Número de Actividades	Precisión obtenida
3	97.98 %
4	97.62 %
5	96.23 %

Se realizaron experimentos entrenando el modelo con 3, 4 y 5 actividades diferentes como caerse al suelo, caminar, beber, hablar por teléfono y comer. Los porcentajes presentados en la Tabla 1 reflejan la precisión del modelo al momento de reconocer la actividad presentada en el video, siendo el 100 el máximo valor a obtener.

Con esto determinamos que el modelo propuesto funcionaría adecuadamente para las cinco actividades usadas. No obstante, esperamos mejorar estos resultados e incrementar la cantidad de actividades que el modelo reconoce para integrarlo a un sistema de monitoreo inteligente completo que sirva de apoyo para el cuidado de los adultos mayores.

Asegurar la protección de datos ayudará a que estos sistemas sean confiables y seguros

Retos y limitaciones

Así como existen diversos modelos para el RAH, también existen varios retos para integrarlos en sistemas complejos para el cuidado de las personas, desde determinar un comportamiento normal o anormal para una persona en específico hasta integrar dichos algoritmos en sistemas que los detecten y se ejecuten acciones acordes a la situación.

A pesar de que las tecnologías de videovigilancia inteligente presentan muchas ventajas para la detección automática de eventos, comportamientos anormales, seguimiento y monitoreo de actividades, también genera polémica. La videovigilancia implica la captura continua de video que para algunas personas puede ser invasiva y por ello surgen inquietudes en el ámbito de la seguridad de datos, privacidad y el uso ético y legal de la información. En el ámbito de la seguridad y privacidad de los datos, a la gente le preocupa quién tiene acceso a su información y lo que se hace con ella. Lo que genera varios retos para el desarrollo de sistemas que garanticen al usuario que el manejo de la información sea seguro cerciorándose del uso ético y legal de sus datos. En cuanto a privacidad, se deben de asegurar la anonimización de los datos, obteniendo el consentimiento de los ocupantes y proveyendo la transparencia sobre la recolección y el uso de dichos datos. Esto sigue siendo un área en el que los investigadores están presentando avances.✷

Agradecimientos

Los autores agradecen a la Secretaria de Educación, Ciencia, Tecnología e Innovación de la Ciudad de México, al Instituto Politécnico Nacional y al Consejo Nacional de Humanidades, Ciencia y Tecnología por el apoyo económico brindado para la realización de esta investigación.

REFERENCIAS

1. Ageing and health. (2022). World Health Organization; World Health Organization: WHO. https://www.who.int/news-room/fact-sheets/detail/ageing-and-health. Recuperado en 25 de Octubre de 2023.

2. Estadísticas a propósito del día internacional de las personas de edad (1o de octubre) datos nacionales. (2019). https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2019/edad2019_Nal.pdf. Recuperado en 25 de Octubre de 2023.

3. Walker-Journey, J. (2022). Best Monitoring Systems For Elderly In The Home. AgingInPlace.org. https://aginginplace.org/medical-alert-systems/monitoring-system-for-elderly-in-the-home/. Recuperado en 27 de Octubre de 2023.

4. Sanchez-Comas A., Synnes K., y Hallberg J. (2020). Hardware for recognition of human activities: A review of smart home and AAL related technologies. *Sensors*, 20(15), 4227.

5. Liu Y.-X., Yang Y., Shi A., Jigang P., y Haowei L. (2019). Intelligent monitoring of indoor surveillance video based on deep learning. En *Proc. 21st International Conference on Advanced Communication Technology* (ICACT). PyeongChang, Korea (South).

6. Zhang Y., Wang C., Wang X., Zeng W., y Liu W. (2021). Fairmot: On the fairness of detection and re-identification in multiple object tracking. *International Journal of Computer Vision*, 129, 3069–3087.

7. Liu Z., Mao H., Wu C.-Y., Feichtenhofer C., Darrell T., y Xie S. (2022). A ConvNet for the 2020s. En *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, LA. pp. 11966-11976.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

SOBRE LOS AUTORES



Elizabeth López Lozada es un estudiante de doctorado en el Laboratorio de Robótica y Mecatrónica del Centro de Investigación en Computación del Instituto Politécnico Nacional. Obtuvo su grado de maestría en Ciencias en Ingeniería de Cómputo en el Centro de Investigación en Computación del Instituto Politécnico Nacional en 2020. Interesada en temas de inteligencia artificial, reconocimiento de patrones, visión por computadora, sistemas embebidos y robótica.



Juan Humberto Sossa Azuela es doctor en Informática por el Instituto Nacional Politécnico de Grenoble, Francia. Es profesor de tiempo completo y Jefe del Laboratorio de Robótica y Mecatrónica del CIC-IPN. Es miembro Emérito del SNII. Es miembro de la Academia Mexicana de Ciencias, de la Academia de Ingeniería, la ACM y Fellow de la Sociedad Mexicana de Inteligencia Artificial. Es también miembro Senior del IEEE y de la INNS. Fue galardonado en 2023 con el Premio a la Investigación en IPN en el área de investigación básica. Sus áreas de investigación son en Inteligencia Artificial,



Elsa Rubio Espino obtuvo el grado de Maestra en Ciencias en Ingeniería Eléctrica en la Especialidad en Control Automático, en el CINVESTAV del IPN y obtuvo el grado de Doctora en Ciencias en la Especialidad de Sistemas Automáticos, por el Instituto Nacional de Ciencias Aplicadas de Toulouse, Francia. Es profesora de tiempo completo en el CIC-IPN. Sus áreas de investigación son referentes a la robótica colaborativa, el reconocimiento de patrones para el monitoreo de actividades humanas; las redes de sensores y la visualización de las bases de datos.



Imagen tomada de:
https://www.techslang.com/the-key-features-of-an-intelligent-monitoring-platform/.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

ARTÍCULO ACEPTADO

Cómputo inteligente en el diseño de celdas de hidrógeno para la generación de energía limpia

Luis Moreno Mendieta, Rolando Menchaca Méndez, Adriana Menchaca Méndez y Marisol Flores Garrido

La Urgencia de Alternativas en la Generación de Energía y el Futuro Prometedor de las Celdas de Combustible

El término *energía limpia* refiere, entre otras cosas, a la electricidad producida por fuentes que no emiten gases de efecto invernadero. Tradicionalmente, los combustibles fósiles han sido la principal fuente de energía, pero son recursos limitados y su uso ha llevado a graves problemas ambientales: contaminación del aire y del agua, destrucción de hábitats naturales y un aumento preocupante en emisiones nocivas, que contribuyen al calentamiento global. Tan solo en 2019 las emisiones de dióxido de carbono relacionadas con la energía alcanzaron un récord de más de 33 gigatoneladas [1]. Es crucial migrar hacia fuentes de energía más limpias y sostenibles para mitigar el cambio climático, reducir el uso de recursos finitos, mejorar la seguridad energética y fomentar un desarrollo económico sostenible. De no actuar, enfrentaremos graves impactos climáticos que amenazarán los ecosistemas, la salud humana y la estabilidad mundial.

Entre las alternativas para generar energía de manera limpia, las celdas de combustible (Figura 1) sobresalen por su alta eficiencia para convertir la energía química de la reacción entre un combustible (hidrógeno, gas natural o biogás) y oxígeno en electricidad.



Figura 1. Un *stack* de celdas de combustible [4].

En el proceso de las celdas de combustible de hidrógeno (H_2), este elemento parte del ánodo (electrodo que libera electrones) y se divide en iones de hidrógeno (H^+) y electrones. Estos últimos fluyen a través de un circuito externo hacia el cátodo (electrodo que recibe electrones), generando corriente eléctrica. Mientras tanto, los iones de hidrógeno pasan a través de una membrana electrolítica hacia el cátodo, donde se combinan con oxígeno (O_2) y con los electrones que han recorrido el circuito externo; esta reacción produce sólo agua y calor como subproductos, completando el proceso de generación de energía eléctrica sin emisiones contaminantes (Figura 2) [2].

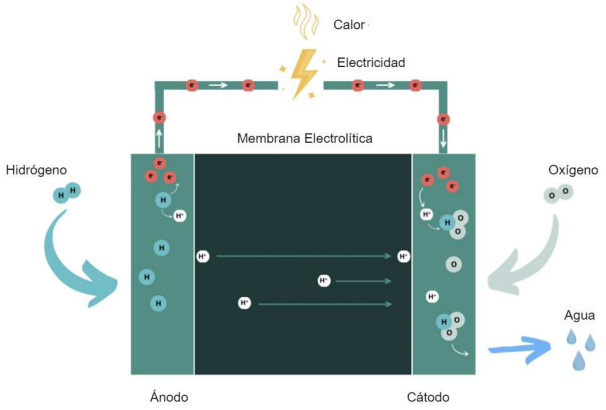


Figura 2. Funcionamiento de una celda de combustible.

Las celdas de combustible de hidrógeno no solo ofrecen una fuente de energía limpia, aprovechando el elemento más abundante en la Tierra, también son más del doble de eficientes que las tecnologías de combustión convencionales. Por ejemplo, una central eléctrica que opera mediante combustión tiene una eficiencia del 33 al 35 %, mientras que las celdas de combustible pueden alcanzar eficiencias de hasta el 60 %, o más si se utiliza cogeneración.

El cómputo inteligente puede ser crucial para mejorar el diseño y acelerar la implementación de celdas de hidrógeno a gran escala.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

Además, son silenciosas, tienen menos partes móviles y son versátiles en sus aplicaciones; incluso, el excedente de energía de fuentes renovables intermitentes (solar, eólica) puede almacenarse como hidrógeno y utilizarse para alimentar la red eléctrica cuando sea necesario [3].

Con tantas ventajas, podría sorprender que la adopción de celdas de combustible sea tan limitada, pero los altos costos de producción y algunos problemas de durabilidad son barreras significativas para su comercialización masiva. Innovar para reducir estos problemas es clave para que las celdas sean una opción competitiva frente a tecnologías tradicionales.

En este artículo describimos una investigación original, realizada en México, orientada al desarrollo de celdas de combustible. Para ello, presentaremos tres desafíos importantes en su construcción y la forma en que el uso de cómputo inteligente, a través de redes neuronales artificiales y computación evolutiva, puede ser crucial para mejorar el diseño y acelerar la implementación de esta tecnología a gran escala.

Desafío 1: Establecer un Modelo para el Diseño de Celdas

Las celdas de combustible no se producen masivamente debido a los altos costos que implican su diseño y construcción. Por un lado, requieren de materiales escasos y de alto costo financiero como el iridio y el platino. Por otro lado, las celdas presentan desafíos importantes en cuanto a diseño. Consideremos, por ejemplo, el manejo del agua. Algunas celdas de combustible usan membranas poliméricas que permiten la conducción de iones (H+) y que dependen del agua para el transporte iónico y para la refrigeración que, a su vez, incide en la durabilidad de los materiales. La presión cerca de las membranas debe ser inferior a la de los reactivos para no impedir su flujo pero, como algunos productos de la reacción generan agua, también se necesita una estrategia para recuperar la que queda atrapada en la salida del ánodo. La cantidad de agua a recuperar depende de las características operativas (temperatura, presión) y del estado de la membrana, y debe lograr la hidratación de esta última pero evitando la inundación del electrodo.

Así, la relación entre las variables que afectan la potencia de salida y la eficiencia de una celda es sumamente compleja. Modelarla involucra no solamente el estudio de las interacciones que existen entre los reactantes (H_2, O_2), sino también con el medio, los materiales de diseño, las geometrías de los componentes, la mecánica del flujo interno y otra serie de fenómenos.

Aunque se puede intuir que podrían usarse matemáticas para optimizar la cantidad de materiales costosos y encontrar la configuración de variables que permite maximizar la potencia de las celdas, optimizar requiere tener una función objetivo y el diseño de las celdas es un caso en el que no se cuenta con una función matemática explícita que relacione todas las variables involucradas.

Afortunadamente, un paradigma que en este momento domina el campo de la inteligencia artificial es el aprendizaje automático. En este enfoque, es posible usar matemáticas para extraer patrones de un conjunto de datos y construir una función que aproxime la relación observada entre las variables que describen instancias del problema y alguna variable de interés.

En particular, las redes neuronales artificiales son modelos computacionales inspirados en el funcionamiento del cerebro humano, diseñados para reconocer patrones a partir de datos. Desde el punto de vista matemático, son estructuras compuestas por nodos (neuronas) interconectados que procesan información mediante la aplicación de funciones matemáticas. Estos nodos se organizan en capas y transmiten información a través de la red, ajustando parámetros de una forma que permite modelar relaciones complejas entre los datos de entrada y salida.

En nuestro caso, fue posible identificar un importante conjunto de datos, obtenidos experimentalmente, que describe la potencia observada en celdas de combustible dependiendo del valor de diez parámetros involucrados en su diseño: la relación de hidrógeno (SH_2) y oxígeno (SO_2) empleadas en la reacción, la temperatura (T), la presión (P), las capacidades de intercambio iónico de la membrana (IEC) y de su capa protectora (IEC_{io}), el grosor de la membrana (δ_{mem}) y de su capa protectora (δ_{io}), la relación de monóxido de carbono respecto de hidrógeno (CO/H_2) y la carga de platino (L_C), que representa el electrolito aplicado para acelerar las reacciones internas [5].

Con estos datos es posible resolver un problema de regresión con algunos modelos tradicionales y con diversas arquitecturas de redes neuronales. En los experimentos que se llevaron a cabo en nuestra investigación, se probó el uso de una red neuronal densa, una red neuronal GRU (*Gated Recurrent Units*) y una red neuronal LSTM (*Long Short-Term Memory*). Estas últimas son un tipo de red neuronal recurrente especializadas en aprender y recordar dependencias a largo plazo en datos secuenciales, gracias a una arquitectura que incluye manejo de memoria y puertas que controlan el flujo de información. Las redes LSTM son las que mostraron mejor desempeño.

Así, a través del uso de inteligencia artificial es posible extraer patrones de datos y construir un modelo paramétrico que, ahora sí, nos permita explorar formas de optimizar la potencia de las celdas de combustible.

Desafío 2: Maximizar Eficiencia

Para hacer que las celdas de combustible sean viables como una tecnología de pronta adopción es necesario optimizar su proceso de producción. Una vez que se cuenta con un modelo paramétrico que permite calcular la po-

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

tencia de salida de la celda a partir de valores de las variables de diseño, es posible integrarlo en una arquitectura de optimización que se encargue de analizar el espacio de búsqueda definido por dichas variables con el propósito de encontrar los valores que maximicen la potencia.

Este problema de optimización puede resolverse utilizando algoritmos evolutivos (AEs), que son métodos computacionales inspirados en principios de evolución natural y genética. En los AEs, los individuos son soluciones candidatas cuya aptitud es mejorada mediante un proceso de evolución simulada que en cada iteración calcula una nueva generación a partir de los individuos que componen a la generación actual. Los AEs son especialmente eficaces para problemas complejos en los que el espacio de soluciones es amplio o no se conoce bien, ya que aprovechan el concepto de supervivencia del más apto para encontrar soluciones óptimas o casi óptimas. Típicamente, la aptitud de un individuo está relacionada con qué tan buena es la solución que representa en términos de los valores de las funciones objetivo.

Utilizando AEs y la red neuronal LSTM para evaluar soluciones candidatas, podemos explorar el espacio de soluciones de nuestro problema de optimización. La Figura 3 ilustra la arquitectura propuesta, donde puede observarse que el AE utiliza a la red neuronal para calcular el valor de la aptitud (*fitness*) de los individuos. En este caso trabajamos con un problema de optimización de objetivo simple (SOOP por su siglas en inglés), pues nos enfocamos en maximizar la potencia de salida de una celda de combustible.

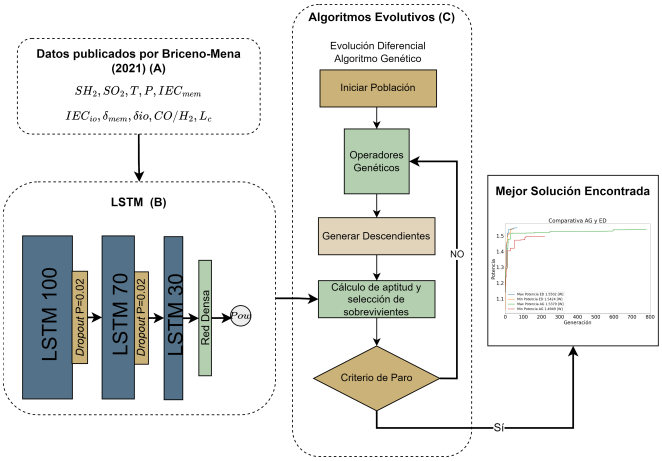


Figura 3. Arquitectura propuesta de un SOOP, en el que se busca maximizar la potencia de una celda de combustible.

Para probar el esquema propuesto, se consideraron dos tipos de AEs: algoritmos genéticos (AG) y evolución diferencial (ED). La principal diferencia entre los AG y ED radica en los operadores genéticos que utilizan; en ED, el operador de mutación genera direcciones de búsqueda a través de diferencias entre soluciones existentes. En los experimentos realizados, ED encontró una mejor solución (1.5447 Watts en 76 generaciones) que AG (1.5379 Watts en 680 generaciones), confirmando lo reportado en la literatura especializada: los algoritmos de ED suelen tener mejor desempeño que los AGs cuando el espacio de búsqueda está conformado por variables continuas.

Un camino a la adopción de celdas de combustible es la optimización de la producción usando modelos paramétricos y algoritmos evolutivos.

Desafío 3: Optimización Simultánea de Potencia y Costo Financiero

Optimizar únicamente la potencia de salida de celdas de combustible conduce a soluciones de alto costo financiero. Un tercer y fundamental desafío es considerar también, en el proceso de diseño, el costo que tiene una configuración propuesta de variables. Ahora queremos maximizar la potencia de las celdas *y* minimizar su costo de producción, dos funciones objetivo que parecen estar en tensión.

Los problemas de optimización multiobjetivo (*MOOP* por sus siglas en inglés) consisten en la optimización simultánea de dos o más funciones que presentan cierto grado de conflicto entre ellas, es decir, si mejoramos una función, tendemos a empeorar la otra. Los algoritmos que resuelven este tipo de problemas se enfocan en identificar el frente de Pareto: un conjunto de soluciones que no se pueden mejorar en un objetivo sin empeorar en al menos algún otro. Para dar un ejemplo en el contexto de nuestro problema, supongamos que se tienen cuatro opciones (*costo*, *potencia*) para el diseño de las celdas: $A = (\$4500, 1{,}42W)$, $B = (\$5000, 1{,}36W)$, $C = (\$9000, 1{,}52W)$, y $D = (\$8500, 1{,}54W)$. Las opciones B y C definitivamente no son las óptimas y como se observa, A es mejor (domina) a B tanto en costo como en potencia, mientras que D domina a C. Por lo tanto, nuestro frente de Pareto sería el conjunto $\mathcal{PF} = \{A, D\}$, que le serviría a un tomador de decisiones para elegir la solución que es más conveniente a sus necesidades [6].

Para la función de costos financieros de nuestro *MOOP* se usaron las variables que corresponden a cuatro de los insumos principales en la producción de la cel-

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

da: SH_2, δ_{mem}, δ_{io} y L_C. Se partió de la suposición de costos marginales, es decir, que al aumentar el volumen de compras se disminuyen los costos unitarios.

El *MOOP* se resolvió empleando cuatro AEs ampliamente utilizados en optimización multi-objetivo: NSGA-II, R-NSGA-II, MOEA/D y SMS-EMOA. Todos ellos utilizan los mismos operadores genéticos, pero difieren en la forma en que seleccionan a los mejores individuos. NSGA-II y R-NSGA-II utilizan el concepto de dominancia de Pareto, mientras que MOEA/D descompone el MOOP en múltiples SOOPs y SMS-EMOA se basa en una métrica conocida como hipervolumen, utilizada para evaluar aproximaciones del frente de Pareto.

El hipervolumen se define como el área cubierta por las soluciones de la aproximación del frente de Pareto y es precisamente la métrica que utilizamos para evaluar cada uno de los algoritmos. La Figura 4 ilustra el concepto de hipervolumen para un problema de minimización. Se observa un punto de referencia r y un conjunto de soluciones $S = \{a, b, c, d, e\}$. El área gris, definida por S y r, denota el hipervolumen de S. Cualquier solución en esta área es dominada por alguna de las soluciones de S. Si agregamos el punto p, la nueva aproximación del frente de Pareto sería $S' = \{a, p, d, e\}$ y el hipervolumen de S' incluiría el área amarilla.

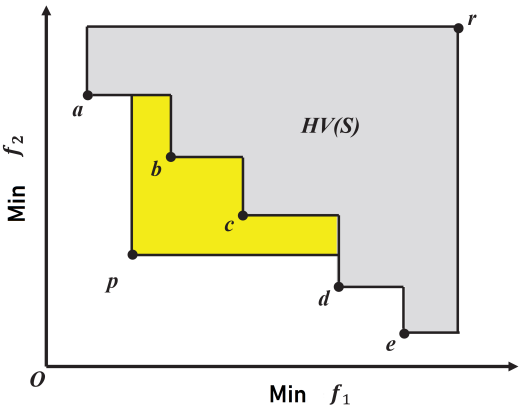


Figura 4. Representación del hipervolumen.

La arquitectura de optimización propuesta se muestra en la Figura 5, donde los AEs utilizan dos funciones objetivo: una red neuronal que calcula la potencia de una celda dados ciertos valores en los parámetros y una función que estima el costo de producción. Una vez obtenidas las aproximaciones del frente de Pareto de cada AE, se combinan en una única aproximación global. La Figura 6 muestra las aproximaciones del frente, normalizadas, encontradas por cada AE. Aunque NSGA-II presenta el mejor comportamiento individual, se observa que los cuatro AEs contribuyen a la construcción del frente de Pareto.

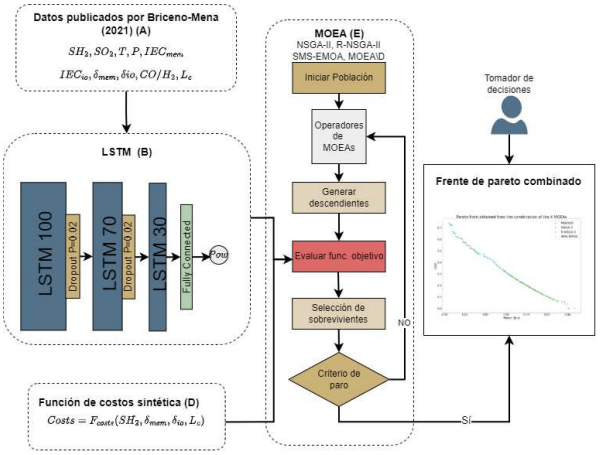


Figura 5. Arquitectura propuesta de un MOOP.

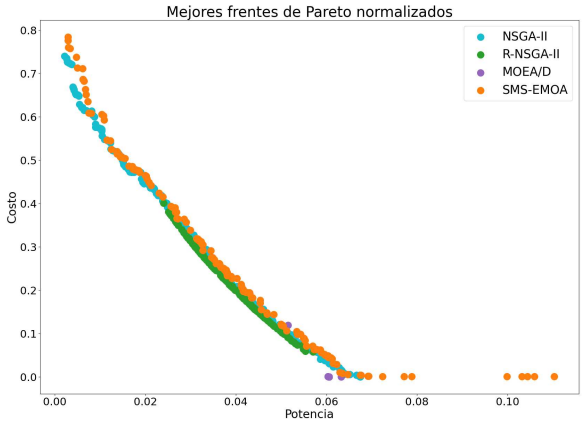


Figura 6. Mejor aproximación del frente de Pareto de cada algoritmo.

La Figura 7 muestra el resultado final de la metodología propuesta y provee soluciones optimizadas (normalizadas), que podrían servir de guía para que un tomador de decisiones pueda seleccionar la que mejor se adecúe a las necesidades particulares de su aplicación.

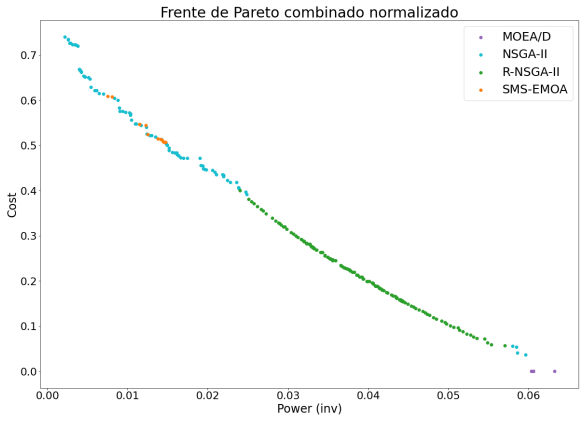


Figura 7. Frente de Pareto combinado.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

Conclusiones

Nos encontramos en un momento clave de la historia de la humanidad. La viabilidad de nuestra forma de vida depende de la manera en que afrontemos el reto que impone el proceso de calentamiento climático global. En la comunidad científica, existe un consenso casi generalizado de que una de las principales causas de este problema es el enorme volumen de emisiones de gases de efecto invernadero producidas por el uso de combustibles fósiles. De ello surge la imperante necesidad de desarrollar tecnologías de generación energética limpias y sustentables que nos permitan alimentar industrias y servicios con un impacto ecológico mínimo.

Las tecnologías de generación de energía limpia traen consigo un conjunto de retos científicos y técnicos que deben abordarse para convertirlas en tecnologías eficientes, económicamente viables y de aplicación generalizada. Enfocándonos en celdas que utilizan hidrógeno como combustible, en este trabajo presentamos un ejemplo del enorme potencial que tiene la IA para contribuir en problemas que, por su importancia, demandan una solución urgente. Consideramos que, a la par del desarrollo tecnológico, será importante desarrollar nuestra imaginación sobre las múltiples y variadas formas en que el cómputo inteligente puede contribuir a construir tecnología que ponga el bienestar y la vida al centro. ✷

REFERENCIAS

1. International Energy Agency (IEA). (2019). Global Energy Review 2019.
2. Blomen, L. J., & Mugerwa, M. N. (Eds.). (2013). Fuel cell systems. *Springer Science & Business Media*.
3. U.S. Department of Energy. (2015). Fuel Cell Technologies Office.
4. SMT LABTECH. (2024). Hydrogen fuel cell. Recuperado el 27 de mayo de 2024, de `https://www.smtlabtech.com/hydrogen-fuel-cell/`.
5. Briceno-Mena, L. A., Venugopalan, G., Romagnoli, J. A., y Arges, C. G. (2021). Machine learning for guiding high-temperature PEM fuel cells with greater power density. *Patterns*, 2(2).
6. Coello, C. A. C., Lamont, G. B., y Van Veldhuizen, D. A. (2007). Evolutionary algorithms for solving multi-objective problems. *New York, Springer*.

SOBRE LOS AUTORES



Luis Moreno Mendieta Es maestro en Ciencias de la Computación, egresado del Centro de Investigación en Computación del Instituto Politécnico Nacional (IPN). Sus líneas de investigación se centran en la investigación de operaciones y optimización, con un enfoque particular en el uso de metaheurísticas y técnicas de aprendizaje automático.



Rolando Menchaca Méndez es Profesor Titular C del Centro de Investigación en Computación del Instituto Politécnico Nacional (CIC-IPN). Doctor en Ciencias egresado de University of California, Santa Cruz. Sus líneas de investigación incluyen teoría de algoritmos, optimización y aprendizaje automático. Es investigador nacional nivel 1.



Adriana Menchaca Méndez es Doctora en Ciencias en Computación por parte del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional. Sus líneas de investigación incluyen optimización mono-objetivo y multi-objetivo utilizando metaheurísticas. Actualmente es Profesora de Tiempo Completo en la Escuela Nacional de Estudios Superiores, Unidad Morelia, de la Universidad Nacional Autónoma de México y es miembro del Sistema Nacional de Investigadoras e Investigadores con nivel 1.



Marisol Flores Garrido es doctora en Ciencias Computacionales por el Instituto Nacional de Astrofísica, Óptica y Electrónica. Trabaja como profesora de tiempo completo en la Escuela Nacional de Estudios Superiores, Unidad Morelia (UNAM) y es miembro del Sistema Nacional de Investigadoras e Investigadores con nivel 1. Sus intereses de investigación se centran en el aprendizaje automático, el análisis de redes y la intersección entre la inteligencia artificial y las humanidades.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

ARTÍCULO ACEPTADO

Aplicaciones de software con inteligencia artificial en el procesamiento de lenguaje natural

Sabino Miranda, Obdulia Pichardo-Lagunas y Bella Martínez-Seis

Introducción

En las últimas décadas, las aplicaciones de software han desempeñado un papel fundamental en la manera en que las personas viven, trabajan y se relacionan. Su importancia radica en su capacidad para facilitar tareas desde cotidianas hasta la optimización de procesos empresariales y mejorar la eficiencia en diversos campos de la ciencia y tecnología. Hoy en día, alguno de sus usos más destacados son: 1) facilitar la comunicación (a través de redes sociales como Facebook, Youtube, WhatsApp, etc.); 2) fomentar la conectividad global (mediante ZOOM, Meet, Teams, entre otros); y, 3) permitir la colaboración en tiempo real y el acceso a recursos y servicios de manera instantánea (por medio de servicios en la nube y de correo electrónico).

Por otro lado, la Inteligencia Artificial (IA) está presente en la vida cotidiana, aunque a veces pase desapercibida, en aplicaciones como los motores de búsqueda (Google, Bing, Baidu), generación automática de texto en videos (Youtube), reconocimiento facial en los teléfonos inteligentes, cálculo de rutas en los mapas (Google Maps), domótica, recomendaciones de productos, asistentes de voz (Alexa, Siri, Asistente de Google, Cortana), etc. Así mismo, la IA ha estado presente en actividades que se consideraban exclusivas para humanos como el jugar al ajedrez, DeepBlue[1] derrotó al campeón Gary Kasparov en 1997; Watson[2] ganó en 2011 el concurso de TV *Jeoparty!*, de preguntas y respuestas que no eran directas o simples hechos; AlphaGo[3] derrotó al campeón del juego de estrategia Go, Lee Sedol, en 2016.

Recientemente, los modelos de lenguaje de gran tamaño (LLMs, por sus siglas en inglés) y sus aplicaciones de software como ChatGPT de OpenAI, Bard (actualmente Gemini) de Google, Llama de Meta, entre otros, han generado gran expectativa en el mundo académico, industrial y en el público en general. Estas aplicaciones han mostrado desempeños sobresalientes en diversos dominios como la generación de texto, generación de texto creativo, resolución a preguntas y respuestas (en matemáticas, ciencia, historia, gramática), traducción, resolución de exámenes, etc. Estas tecnologías no solo han ampliado las capacidades de las aplicaciones convencionales, sino que también han introducido nuevas formas de comprensión y procesamiento de datos. Lo anterior se ha dado a una escala antes inimaginable abriendo nuevas posibilidades para la innovación y la transformación digital.

Aplicaciones de la inteligencia artificial

La IA tiene por objeto la comprensión y la construcción de entidades inteligentes, esto es, máquinas que puedan calcular cómo actuar de manera segura y efectiva en situaciones nuevas. Actualmente, abarca una gran variedad de subcampos, entre ellos se incluyen tanto áreas de propósito general (p. ej., el aprendizaje, el razonamiento y la percepción) como específicas (p. ej., el ajedrez, la demostración de teoremas, el manejo de automóviles, o el diagnóstico de enfermedades) [1].

Entre las múltiples ramas de la IA, la visión artificial y el procesamiento del lenguaje natural (PLN) se destacan como áreas cruciales que han experimentado avances significativos en los últimos años.

En el ámbito de la visión artificial, las máquinas han adquirido la habilidad de interpretar y comprender el mundo visual de manera similar a los seres humanos. Esto implica el desarrollo de algoritmos y modelos capaces de analizar imágenes y videos para reconocer patrones, objetos y contextos. Aplicaciones prácticas de la visión artificial incluyen sistemas de reconocimiento facial, vehículos autónomos, diagnóstico médico basado en imágenes, generación de imágenes basada en texto (p. ej., DALL-E[4] , Midjourney[5] y Stable Difussion[6]) y mucho más.

Por otro lado, el PLN es una rama de la IA que se centra en permitir a las máquinas entender, interpretar y generar lenguaje humano de manera coherente. Con el PLN, las aplicaciones pueden comprender el contexto de las conversaciones, comprender los sentimientos en el texto, traducir idiomas de manera automática, generar texto de manera creativa y proporcionar respuestas contextuales en interacciones con el usuario. Esta capacidad ha impulsado la creación de asistentes virtuales cada vez

[1] https://www.ibm.com/ibm/history/exhibits/vintage/vintage_4506VV1001.html
[2] https://www.ibm.com/support/pages/what-watson-ibm-takes-jeopardy
[3] https://deepmind.google/technologies/alphago
[4] https://openai.com/dall-e-2
[5] https://www.midjourney.com
[6] https://github.com/Stability-AI/generative-models

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

más inteligentes, comprensión de las preferencias de los usuarios por medio del perfilado, y herramientas que facilitan la comunicación hombre-máquina de manera más natural y efectiva.

Actualmente, el campo de la IA es muy amplio y su domino de aplicación diverso, por lo que, en las siguientes secciones nos enfocamos en tres aplicaciones y tareas relacionadas al PLN: los sistemas conversacionales, los asistentes de voz, y el análisis de sentimientos. Estas aplicaciones son interesantes debido a sus retos tanto académicos como ingenieriles, así como su popularidad.

Sistemas conversacionales y modelos de lenguaje de gran tamaño

Uno de los primeros sistemas conversacionales o *chatbots* fue ELIZA [2]. ELIZA simulaba tener conversaciones escritas con el usuario en lenguaje natural usando la puntuación normal y estructuras de las oraciones. ELIZA utilizaba los patrones conversacionales del usuario, reglas de transformación, entre otros mecanismos para responder, dando la ilusión que comprendía la conversación. Actualmente, los avances en investigación en redes neuronales artificiales y la creación de la arquitectura de transformadores (*Transformers*) [3] han impactado ampliamente en el aprendizaje de máquina y en el área del PLN. Esta arquitectura se ha aplicado a múltiples dominios y es parte fundamental de sistemas como ChatGPT de OpenAI (ver Figura 1) y Gemini de Google. Éstos son sistemas con los que el usuario interactúa por medio de instrucciones o *prompts*, obteniendo respuestas de parte de ellos. Así mismo, la arquitectura está presente en los modelos de lenguaje de gran tamaño (LLMs, por sus siglas en inglés) como GPT[7] de OpenAI, PaLM-2[8] de Google, Llama[9] (de código abierto) de Meta, entre muchos más.

Esta familia de modelos de lenguaje de gran tamaño se entrenan con enormes cantidades de datos (de millones a billones de tokens —un token representa una secuencia de símbolos equivalente a palabras—). A la fecha, han demostrado obtener desempeños sobresalientes en diferentes tareas como la comprensión del lenguaje, generación de textos, dar respuestas a preguntas, traducción entre idiomas, mantener diálogos, resolución de exámenes, etc.

Tanto ChatGPT como Gemini son aplicaciones propietarias de las cuales no se conocen detalles de su implementación. El modelo GPT-3 cuenta con 175 mil millones de parámetros y GPT-4 se estima en 1 billón de parámetros [4]. PaLM cuenta con 540 mil millones de parámetros y PalM-2 se estima que es menor en el número de parámetros que PaLM pero más eficiente [5]; el número de parámetros indica el tamaño del modelo. Estos modelos de lenguaje propietarios están fuertemente ajustados con técnicas de aprendizaje por refuerzo a las preferencias humanas. Esto permite que la generación de texto, resolución de exámenes, preguntas y respuestas o diálogos tengan un buen desempeño, usabilidad y seguridad. Este proceso requiere de costos significativos en cómputo y anotadores humanos.

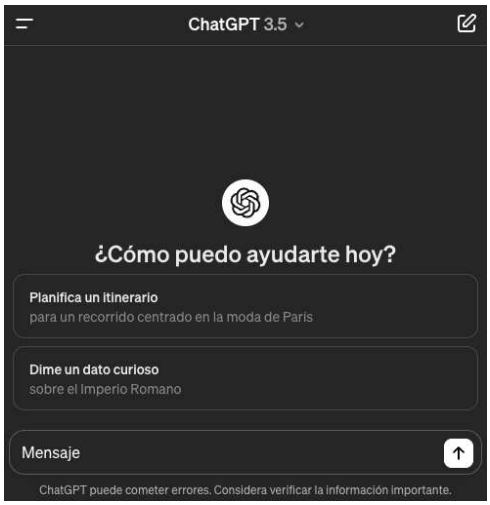


Figura 1. Interfaz de chatGPT.

Por otro lado, la familia Llama son modelos de lenguaje grandes preentrenados y ajustados, son de código abierto y están disponibles para la comunidad (tanto los modelos preentrenados como el código fuente) para fines de experimentación y comerciales hasta cierto punto. Llama 2 y Llama 2-chat cuentan con hasta 70 mil millones de parámetros. Llama 2-chat está optimizado para diálogos y en las evaluaciones del sistema se alcanzan desempeños a la par de los modelos propietarios [6].

Para este tipo de familia de modelos de lenguaje de gran tamaño, se han llevado a cabo experimentos donde se muestra que mientras más grande es el modelo y la cantidad de datos de entrenamiento, es mejor el desempeño en diversas tareas como las antes mencionadas. Sin embargo, esto puede cambiar según los resultados presentados para el modelo PaLM-2 ya que es un modelo de menor tamaño que PaLM y su desempeño es superior [5].

Un problema que se ha identificado en este tipo de modelos es que sufren de "alucinaciones" o sesgos sociales en la información generada. Es decir, en ocasiones proporcionan información incorrecta de hechos o fechas, tornan racistas las respuestas, o generan información errónea o fantasiosa. En sí, esta área sigue siendo investigada y se siguen proponiendo diversas soluciones tanto en la

[7] https://platform.openai.com/docs/models
[8] https://ai.google/discover/palm2/
[9] https://llama.meta.com

forma de entrenar como en la manera de evitar estos fallos y sesgos latentes en estos modelos de lenguaje.

Sin duda, este tipo de chatbots inteligentes está revolucionando la forma en que se busca información, se aprende, se hace investigación, se descubre conocimiento, además de otras maneras de aprovechar esta tecnología en las actividades cotidianas de las personas.

Asistentes de voz y reconocimiento de voz

Uno de los objetivos de la IA es la realización de diálogos naturales entre los humanos y las máquinas. Recientemente, los asistentes de voz se han convertido en parte de la vida cotidiana, puesto que están presentes en diversos dispositivos como teléfonos inteligentes, tabletas, computadoras, bocinas inteligentes, etc. Muchas compañías han usado tecnologías de sistemas de diálogo para implementar varios asistentes de voz como Amazon Alexa, Siri de Apple, Asistente de Google, Cortana de Microsoft. En general, el asistente de voz convierte la voz en texto para posteriormente usarlo y recolectar la información de diferentes fuentes y dar respuesta al usuario (ver asistente de voz que se muestra en la Figura 2). Actualmente, este tipo de asistentes de voz proporcionan un conjunto similar de servicios, se usan como asistentes de preguntas y respuestas, y comandos de voz para ejecutar alguna orden; ya que no pueden mantener una conversación prolongada, solo unas pocas interacciones. En el fondo, para realizar el trabajo primero se hace el reconocimiento de la voz, identificando la palabra de activación, por ejemplo, "*Oye, siri*", "*Alexa*", "*Ok, Google*"; en seguida se captura la intención del usuario (ya sea la pregunta o comando que se desea ejecutar); y, finalmente, se da la respuesta mediante voz. Para estos sistemas, el procesamiento de la voz se realiza en los servicios en la nube del proveedor.



Figura 2. Asistentes de voz (Imagen generada con DALLE-2.)

Una parte importante del proceso es el reconocimiento de voz que consiste en cuatro módulos: preprocesamiento, extracción de características, modelo de clasificación y modelo de lenguaje. En el preprocesamiento, se prepara la señal de voz y se reduce el ruido de la señal de voz por medio de preénfasis, normalización, muestreo, ventaneo, etc. En la extracción de características, se crean los vectores de características del habla típicamente se usan MFCC, PLP, LPC, etc. El modelo de clasificación es el que se encarga de predecir el texto correspondiente basado en el vector de características extraídas. Tradicionalmente, se han usado modelos ocultos de Markov, modelos de mixturas gaussianas, y, recientemente, modelos de redes neuronales. El modelo de lenguaje se usa para reconocer los fonemas predichos, usa reglas y semántica del lenguaje; sistemas modernos de reconocimiento del habla prescinden de este módulo [7].

Por ejemplo, en el caso de Siri, para detectar la palabra de activación se usa un modelo basado en una red neuronal prealimentada con cinco capas ocultas completamente conectadas, del mismo tamaño, que pueden variar entre 32, 128 o 192 neuronas, dependiendo de la memoria y de restricciones de energía del dispositivo, al final de las predicciones de esta red, una red neuronal recurrente integra las predicciones de todos los cuadros (*frames*) que se generaron a partir de la señal de entrada [8].

Por otro lado, también existen asistentes de voz de código abierto como Mycroft[10], lo que permite experimentar y explorar la implementación de nuevas características y entrenar modelos propios o de un dominio específico.

Sin duda, los asistentes de voz cada vez más se integrarán a la vida cotidiana, ya que el lenguaje hablado es una interfaz natural para las personas.

Análisis de sentimientos y perfilado de usuarios

En los últimos años, las actividades industriales alrededor del análisis de sentimientos y el perfilado de usuarios han prosperado significativamente. Así mismo, un gran número de nuevas empresas han surgido, y los sistemas de análisis de sentimientos asociados al sentir de los usuarios han encontrado aplicación en casi cualquier dominio social y de negocio. La información que las personas escriben en las distintas plataformas sociales (Twitter/X, Facebook, Amazon (reseñas), etc.) es útil para las empresas, gobiernos e individuos que quieren beneficiarse de los comentarios que se publican sobre sus actividades o productos. En particular, el perfilado de usuarios es el proceso de analizar datos sobre los usuarios y entender sus intereses, demográficos (edad, género, etc.) y comportamientos. Esta información tiene muchas aplicaciones, como el modelado de usuarios para mejorar la experiencia del usuario, publicidad dirigida, detección de

[10] https://github.com/MycroftAI/mycroft-core

actividad maliciosa, medición de las preferencias de los usuarios sobre fenómenos sociales, por ejemplo, las preferencias políticas, el discurso de odio hacia migrantes o mujeres, posturas ante movimientos sociales, entre otros [9].

El análisis de sentimientos y el perfilado de usuarios se ha abordado usualmente como un problema de clasificación de textos. Se ha impulsado el desarrollo de estas áreas por medio de competencias internacionales especializadas en diversos tipos de clasificación de texto. El objetivo de estas competencias es impulsar el desarrollo de software de IA y de nuevas técnicas de IA que propongan resolver estos campos. Por ejemplo, se cuenta con competencias sobre la detección de polaridad, identificación de misoginia, identificación de agresividad, detección de preferencias de ideología política, detección de emociones (enojo, miedo, alegría, tristeza), entre otros.

Para modelar los textos en el análisis de sentimiento y el perfilado de usuarios; usualmente, se usa un modelado vectorial. Los métodos tradicionales, como los modelos de bolsa de palabras, pueden extraer características de los textos. Estas técnicas se basan en características léxicas que deben determinarse a través de algún esquema de ponderación basado en la frecuencia del término (p. ej., TF, TF-IDF) o basado en entropía, con los cuales se entrena el clasificador. Por otro lado, en los últimos años, se han utilizado técnicas de aprendizaje profundo para aprender modelos de lenguaje. La estrategia general utiliza modelos de lenguaje previamente entrenados para obtener vectores densos (*embeddings*) de alta dimensión con una arquitectura de *Transformers*. Por ejemplo, para el idioma inglés, los modelos BERT [10] y sus variantes como RoBERTa; en el caso del idioma español, BETO, MarIA, entre otros basados en modelos BERT.

Por mencionar un caso, está el mejor sistema evaluado para la detección de ideología política en la competencia PoliticES-2023[11]. Éste usa una arquitectura de redes neuronales profundas que combina modelos preentrenados basados en una arquitectura de *Transformers*, BETO y MarIA. Debido a que la arquitectura de *Transformers* usada tiene limitaciones de longitud máxima, se divide el texto en partes y se usan un esquema de votación para las predicciones parciales y determinar la clase final.

Otro caso sería el mejor sistema evaluado en la detección de odio contra mujeres e inmigrantes para el idioma inglés en la competencia HatEval 2019[12]. Este sistema utiliza una representación del texto basada en *Universal Sentence Encoder* que codifica el texto en un vector denso; este vector es la entrada a un clasificador de Máquinas de Vectores de Soporte (SVM) con kernel RBF que calcula la predicción final.

[11] https://sites.google.com/view/iberlef-2023/tasks
[12] https://competitions.codalab.org/competitions/19935

Conclusiones

Las aplicaciones de software con IA, particularmente los asistentes de voz, tienen un futuro promisorio como integradores de tecnologías desarrolladas en diversas áreas del PLN (p. ej., el análisis de sentimientos y emociones, los mecanismos desarrollados para los chatbots, entre otras). No pasará mucho tiempo en que veamos la integración de varias de estas tecnologías en los asistentes de voz cerrando la brecha entre la ciencia ficción y la realidad cotidiana. A final de cuentas, toda las tecnologías desarrolladas tienen un fin primordial, asistir al humano en sus tareas para que pueda tomar las mejores decisiones basadas en el conocimiento que les puedan proporcionar. La IA está demostrando ser una tecnología que cambiará radicalmente el mundo donde vivimos. A medida que se continúa explorando nuevas fronteras de la IA, es imperativo mantener un enfoque ético que pueda garantizar que la IA beneficie a la humanidad en su conjunto.✷

REFERENCIAS

1. Russel, S. y Norving, P. (2020). Artificial Intelligence: A Modern Approach. 4th. Hoboken: Pearson.
2. Weizenbaum, J. (1966). ELIZA—a Computer Program for the Study of Natural Language Communication between Man and Machine. *Commun. ACM*, 9(1), 36–45.
3. Vaswani, A. Shazeer, N., Parmar, N. et al. (2017). Attention is All You Need. En *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach, California, USA: Curran Associates Inc., 6000–6010. ISBN 9781510860964.
4. OPENAI. GPT-4 Technical Report. 2023. [Recuperado: 2024-04-10] doi: https://doi.org/10.48550/arXiv.2303.08774.
5. Anil, R., Dai, A., Firat, O. et al. (2023). PaLM 2 Technical Report. [Recuperado: 2024-04-10] doi: https://arxiv.org/abs/2305.10403.
6. Touvron, H., Louis, M., Stone, K. et al. (2023). Llama 2: Open Foundation and Fine-Tuned Chat Models. [Recuperado: 2024-04-10] doi: https://doi.org/10.48550/arXiv.2307.09288.
7. Malik, M., Malik, M. K., Mehmood, K. et al. (2021). Automatic Speech Recognition: a survey. *Multimedia Tools and Applications*, 80(6), 9411–9457. ISSN 1573-7721
8. APPLE, Siri Team (2017). Hey Siri: An On-device DNN-powered Voice Trigger for Apple's Personal Assistant. [Recuperado 23-11-2023] https://machinelearning.apple.com/research/hey-siri.
9. Bevendorff, J., Chulvi, B., Fersini, E. et al. (2022). Overview of PAN 2022: Authorship Verification, Profiling Irony and Stereotype Spreaders, Style Change Detection, and Trigger Detection. En Hagen, M., Verberne, S., MacDonald, C. et al. (eds.). Advances in Information Retrieval. Cham: Springer, pp. 331–338. ISBN 978-3-030-99739-7.
10. Devlin, J., Chang, M., Lee, K. et al. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. En *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1, 4171–4186.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

SOBRE LOS AUTORES



Sabino Miranda recibió el grado de Doctor en Ciencias de la Computación del Centro de Investigación en Computación del Instituto Politécnico Nacional (CIC-IPN). Actualmente es Profesor en la Universidad Autónoma de la Ciudad de México (UACM). Sus intereses de investigación incluyen el procesamiento del lenguaje natural, el aprendizaje automático y la clasificación de textos.



Obdulia Pichardo-Lagunas recibió los grados de doctorado y maestría en Ciencias de la Computación del Centro de Investigación en Computación del Instituto Politécnico Nacional (CIC-IPN). Actualmente es Profesora Titular de la Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas (UPIITA-IPN). Sus intereses de investigación incluyen el procesamiento del lenguaje natural.



Bella Martínez-Seis recibió el grado de Doctor en Ciencias de la Computación del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN). Tiene una Maestría en Ciencias de la Computación del Centro de Investigación en Computación (CIC-IPN). Actualmente es Profesora Titular de la Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas (UPIITA-IPN). Sus intereses de investigación incluyen Minería de Datos y Análisis de Redes Sociales.

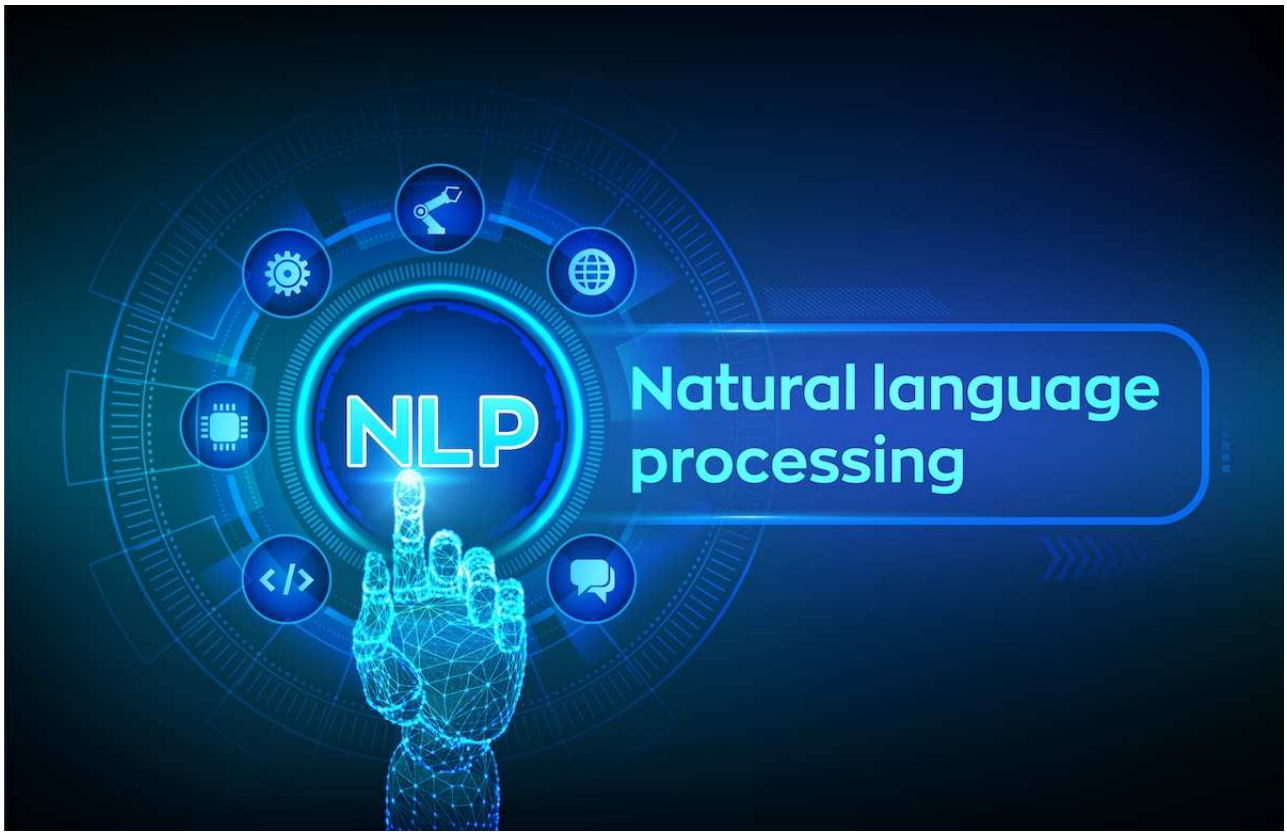


Imagen tomada de: https://iasolver.es/procesamiento-del-lenguaje-natural/.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

ARTÍCULO ACEPTADO

Comparación de dos modelos de redes neuronales preentrenados aplicados en la detección de plagas en cultivos de aguacate

Jesús García Palomera, Jesús Alberto Verduzco Ramírez, Juan García Virgen y Luis Eduardo Morán López

México es el principal productor mundial de aguacate, con una aportación aproximada del 46 % del total de las exportaciones mundiales. La cadena productiva de este fruto es afectada por plagas, enfermedades, entre otros factores, repercutiendo negativamente en la producción. Una de las plagas con mayor presencia en los cultivos de aguacate es la araña cristalina (Oligonychus perseae). Este ácaro provoca afectaciones en la epidermis de las hojas, dando origen a puntos de tejido muerto que afecta la fotosíntesis y causa la caída del follaje, dejando expuestos los frutos a la intemperie, provocando su deshidratación y muerte.

La detección temprana de esta plaga es crucial para controlar su propagación y mitigar sus efectos en el cultivo. Los métodos basados en inspección visual realizada por personal son costosos, laboriosos y propensos a errores. Una alternativa para optimizar este proceso, son las técnicas de visión artificial basadas en modelos de redes neuronales.

Este trabajo reporta los resultados de la evaluación de los modelos de redes neuronales YOLOv5 y MobileNetV3 en la detección de la plaga mencionada. Para este fin, se utilizó un conjunto de 180 imágenes públicas, el 92 % de las cuales se destinaron para el entrenamiento, y el 8 % para la validación. Los criterios utilizados en la evaluación son la exactitud de la clasificación, y el consumo energético invertido durante el entrenamiento.

Los resultados indican que YOLOv5 obtuvo un mejor desempeño, logrando una exactitud del 93 % en la clasificación de imágenes, en comparación con el modelo MobileNetV3, que alcanzó 73 %. Sin embargo, MobileNetV3 demostró un mejor rendimiento en términos de consumo energético.

Introducción

El aguacate es uno de los productos más exitosos de la exportación agroalimentaria nacional, con una aportación de casi la mitad del valor de las exportaciones mundiales. De acuerdo con información oficial, se estima que para el 2030 la producción de aguacate alcanzará 3.16 millones de toneladas, con un crecimiento promedio anual del periodo 2016-2030 del 3.49 %. El valor de las exportaciones estimadas para 2030 ascendería a 4,655 millones de dólares [1]. Derivado del crecimiento en la producción de aguacate, las superficies sembradas y cosechadas en el país han aumentado, alcanzando un total de 252,132 hectáreas (ha) sembradas en 2022 [2], como se ilustra en la Figura 1.

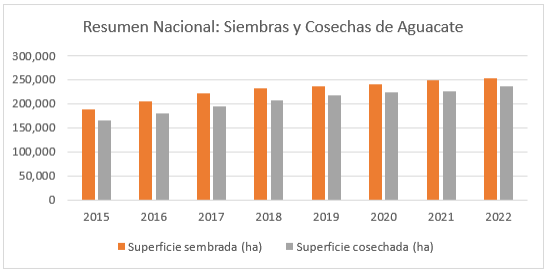


Figura 1. Resumen Nacional de Siembras y Cosechas de Aguacate. Fuente: Elaboración propia con datos del SIAP [2].

Con el incremento en la producción de aguacate y a pesar de los esfuerzos para monitorear y detectar plagas y enfermedades en los cultivos por medio de personal capacitado, aún existen pérdidas en la producción debido a la presencia de estos factores. Las pérdidas económicas derivadas de estos problemas impactan negativamente la cadena de producción de este fruto en el país [1].

La detección oportuna de plagas asociadas con cultivos es una preocupación constante en la agricultura. Uno de los métodos novedosos para optimizar la detección de plagas es el procesamiento de imágenes mediante técnicas de visión computacional, que utiliza modelos de redes neuronales y aprendizaje automático con el fin de detectar y clasificar patrones en las imágenes. Existen diferentes arquitecturas y modelos de redes neuronales que se utilizan actualmente en la detección de patógenos en la agricultura, dos de los modelos preentrenados con mayor aplicación son YOLOv5 y MobileNetV3.

YOLOv5 (You Only Look Once) [3] es un modelo de red neuronal convolucional usado para la detección y clasificación de objetos. YOLOv5 utiliza un enfoque de detección basado en problemas de regresión, escaneando la imagen completa durante el tiempo de prueba para codificar tanto la información contextual como la apariencia de los objetos. Por su parte, MobileNetV3 [4]

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

es un modelo de red neuronal que factoriza eficazmente las convoluciones tradicionales al dividirlas en pasos más sencillos. Al hacer esto, MobileNetV3 facilita la generación de características, como si se tratara de ensamblar un rompecabezas al encajar piezas más sencillas y simples.

A continuación, se resumen algunos trabajos de investigación que han evidenciado la efectividad del uso de modelos de redes neuronales para la detección de patógenos en cultivos.

Zhang, L. et al [5], presentaron un algoritmo de aprendizaje profundo basado en YOLOv5, el cual tiene el objetivo de asistir a los agricultores en el diagnóstico de 102 tipos de plagas. Para lograr esto, se propuso integrar un nuevo módulo inspirado en la red MobileNetV3. Los resultados indican que el algoritmo propuesto tiene un mejor desempeño que el original en un conjunto de datos públicos de 102 diferentes plagas. Específicamente, la exactitud de detección fue 2.4 % mayor que la del modelo original.

Khalid, M. et al. [6] aplicaron el modelo entrenado YOLOv5 para detectar manchas en las hojas. Para este propósito, crearon un conjunto de datos con 2,311 imágenes de hojas de árbol de jade. Estas imágenes fueron divididas en dos categorías principales: (1) saludables y (2) no saludables. El modelo entrenado alcanzó un 93 % de precisión.

Por otro lado, Adnan, M. et al. [7] presentaron una propuesta similar al implementar un sistema utilizado para clasificar y detectar enfermedades en las plantas. El conjunto de datos contiene 20,636 imágenes de plantas y sus enfermedades. El sistema propuesto utiliza redes neuronales convolucionales para clasificar las plantas en 12 categorías de enfermedades. Como resultado, obtuvieron una precisión del 98.02 % en las pruebas aplicadas al conjunto de datos utilizado.

En resumen, los trabajos analizados tienen en común la característica de requerir una gran cantidad de imágenes para el entrenamiento de los modelos, así como de recursos computacionales y de energía para el procesamiento y clasificación de patógenos. Estos tres factores incrementan el costo y son de difícil acceso para los agricultores de pequeña escala, quienes disponen de pocos recursos económicos. Además, no ofrecen una solución que pueda utilizarse en campo, por ejemplo, en un dispositivo móvil.

El objetivo de este trabajo es comparar los modelos YOLOv5 y MobileNetV3 en la detección de la plaga araña cristalina (Oligonychus perseae) en cultivos de aguacate utilizando imágenes públicas, y de esta forma, determinar el modelo con mejor desempeño. La solución propuesta está dirigida a productores con reducida capacidad económica que requieren herramientas eficientes y de fácil uso para la detección oportuna de plagas. Se pretende utilizar los resultados obtenidos como parte de un proyecto que permita optimizar el monitoreo y detección de plagas directamente en campos de cultivo. Este trabajo está organizado en secciones que describen la metodología utilizada, los resultados obtenidos, las conclusiones y, por último, algunas propuestas para trabajos futuros.

Metodología y herramientas

Las herramientas utilizadas para el desarrollo de este trabajo fueron YOLOv5 v7.0, MobileNetV3-Large, repositorio de imágenes públicas Roboflow [8], Google colaboratory configurado con GPU T4 [9].

Con el fin de organizar las actividades del proyecto, se utilizaron las pautas del proceso que se ilustran en la Figura 2.

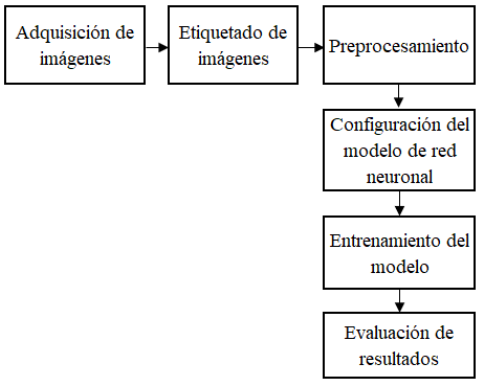


Figura 2. . Proceso utilizado para el desarrollo del trabajo. Fuente: Elaboración propia.

Adquisición y etiquetado de imágenes

Se recolectó un conjunto de 70 imágenes disponibles públicamente, de las cuales 35 contienen hojas de cultivos de aguacate con presencia de araña cristalina y 35 con hojas saludables. Para este propósito, se utilizó la herramienta Roboflow para el etiquetado de las imágenes.

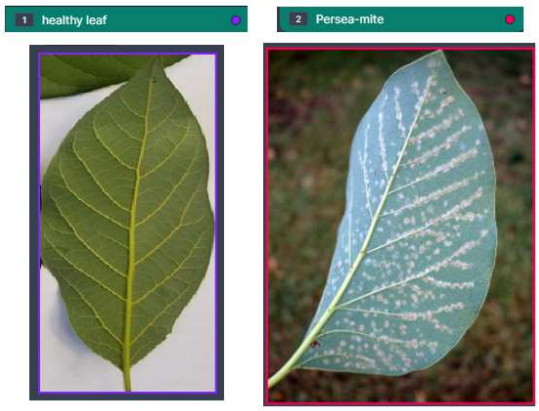


Figura 3. Selección y etiquetado de imágenes saludables y con presencia de araña cristalina utilizando Roboflow.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

Preprocesamiento

Posteriormente, se aplicó el preprocesamiento y el redimensionamiento a un arreglo de 224 x 224 píxeles, generando como resultado un total de 180 imágenes. Los parámetros utilizados se muestran en la Figura 4.

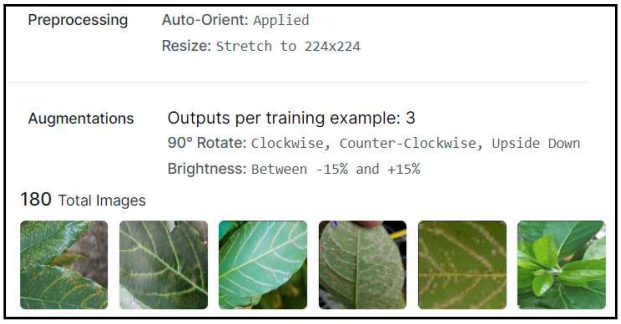


Figura 4. Preprocesamiento y redimensionamiento de imágenes utilizando Roboflow.

Configuración de los modelos

Para homogenizar los criterios de comparación, ambos modelos se configuraron con valores de parámetros similares y el mismo conjunto de imágenes, como se muestra en la Tabla 1.

Tabla 1. Resultados de las pruebas y parámetros de entrenamiento.

Nombre del modelo	YOLOv5	MobileNetV3
Conjunto de datos	Imágenes públicas	Imágenes públicas
Imágenes de entrenamiento	165	165
Imágenes de prueba	15	15
Tamaño de imágen (pixeles)	224 x 224	224 x 224
Ciclos de entrenamiento	100	100
Tasa de aprendizaje (α)	0.001	0.001
Lote de entrenamiento	64	64
Exactitud	93%	73%

Entrenamiento

Una vez generado el conjunto de imágenes y establecida la configuración de los modelos, se procedió a realizar el entrenamiento utilizando la plataforma Google Colaboratory, configurando el GPU T4 como dispositivo de procesamiento. El conjunto de imágenes fue importado desde el repositorio de Roboflow*. Las Figuras 5 y 6 muestran los últimos cinco ciclos del proceso de entrenamiento, así como los valores de exactitud obtenidos.

Epoch	GPU mem	train loss	test loss	top1 acc
96/100	1.25G	0.29	0.27	1
97/100	1.25G	0.306	0.278	0.933
98/100	1.25G	0.294	0.279	0.933
99/100	1.25G	0.358	0.281	0.933
100/100	1.25G	0.313	0.28	0.933

Figura 5. Resultados del entrenamiento del modelo YOLOv5.

```
Epoch 96/100
3/3 [==========] - val_loss: 1092.6808 - val_accuracy: 0.7333
Epoch 97/100
3/3 [==========] - val_loss: 1091.8267 - val_accuracy: 0.7333
Epoch 98/100
3/3 [==========] - val_loss: 1091.6603 - val_accuracy: 0.7333
Epoch 99/100
3/3 [==========] - val_loss: 1091.5359 - val_accuracy: 0.7333
Epoch 100/100
3/3 [==========] - val_loss: 1091.4401 - val_accuracy: 0.7333
```

Figura 6. Resultados de entrenamiento con el modelo MobileNetV3.

Discusión de resultados

Para evaluar el comportamiento de los modelos, se tomaron en cuenta los valores de exactitud de la clasificación. Los resultados obtenidos se resumen en la Tabla 1.

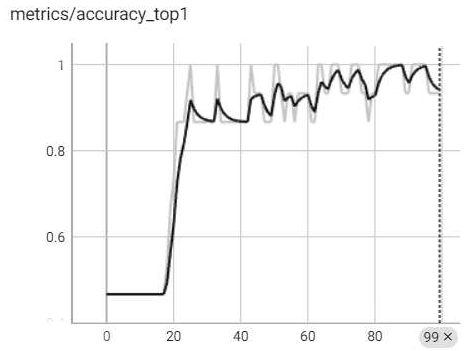
La Figura 7 ilustra el comportamiento de la exactitud para la clasificación de los modelos YOLOv5 y MobileNetV3 en 100 ciclos de entrenamiento. Como se muestra, el valor de exactitud aumenta conforme se incrementan los ciclos de entrenamiento, llegando a un valor máximo de 100 % para YOLOv5 entre los ciclos 60 y 100. Al final del entrenamiento, el valor se estableció en 93 %. En el caso del modelo MobileNetV3, la exactitud de la clasificación se estableció en un 73 % después de la iteración 30 de entrenamiento.

Consumo energético

Uno de los intereses de este proyecto consistió en determinar el consumo energético de los modelos durante el entrenamiento, ya que se busca que la solución sea sustentable y portable a un dispositivo móvil. Para realizar las mediciones de consumo energético, se utilizó como plataforma de experimentación una computadora equipada con sistema operativo Windows 11, 8GB de RAM, procesador AMD Athlon Silver 3050U @2.30GHz y el software HWiNFO64 [10]. La Figura 8 muestra los resultados obtenidos en 100 ciclos de entrenamiento del modelo YOLOv5. Como se observa, el tiempo de entrenamiento fue de 33 minutos 43 segundos, y el consumo de energía promedio fue de 7.9 W.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

1. YOLOv5



2. MobileNetV3

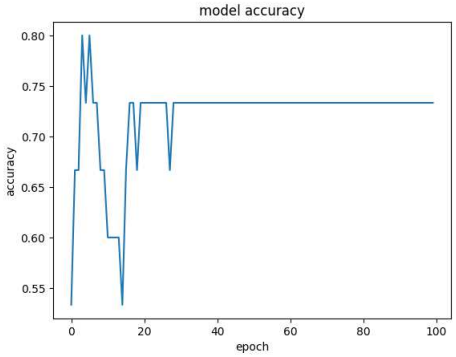


Figura 7. Exactitud de clasificación del modelo YOLOv5(1) vs MobileNetV3(2) en 100 ciclos de entrenamiento.

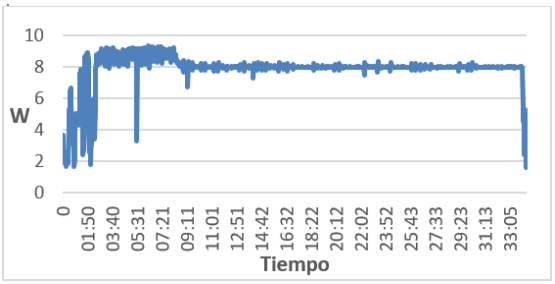


Figura 8. Modelo entrenado YOLOv5: Consumo total de energía del procesador [W].

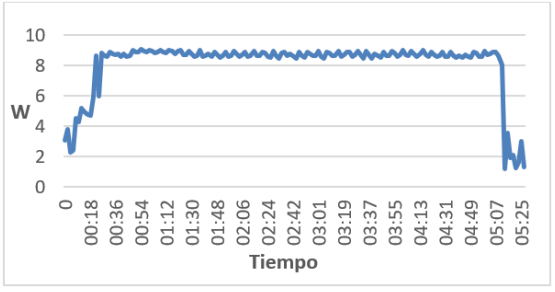


Figura 9. Modelo entrenado MobileNetV3: Consumo total de energía del procesador [W].

La Figura 9 ilustra el consumo de energía del modelo MobileNetV3. Los resultados muestran que el tiempo de entrenamiento fue de 5 minutos 2 segundos, y el promedio de potencia consumida de 8.2 W.

Conclusiones

Este estudio reporta la comparación de la exactitud en la detección de la plaga araña cristalina y el consumo energético durante el entrenamiento de los modelos de redes neuronales YOLOv5 y MobileNetV3. YOLOv5 alcanzó una precisión del 93 % en la clasificación, mientras que MobileNetV3 logró el 73 %. Con respecto al consumo de energía, el de YOLOv5 fue seis veces mayor que el de MobileNetV3.

Estos hallazgos indican que los modelos de redes neuronales pueden ser herramientas efectivas para la detección de plagas en los cultivos de aguacate. Trabajos futuros estarán orientados a medir el comportamiento de ambos modelos utilizando imágenes obtenidas directamente de los campos de cultivo.✷

REFERENCIAS

1. Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación. (2017). Planeación Agrícola Nacional 2017-2030. Accedido: 6 de octubre de 2023. [En línea]. Disponible en: https://www.gob.mx/agricultura/documentos/planeacion-agricola-nacional-2017-2030?state=published.

2. Servicio de Información Agroalimentaria y Pesquera. (2023). Producción Agrícola Nacional. Accedido: 4 de septiembre de 2023. [En línea]. Disponible en: https://www.gob.mx/siap/acciones-y-programas/produccion-agricola-33119.

3. Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2015). You Only Look Once: Unified, Real-Time Object Detection. Accedido: 23 de septiembre de 2023. [En línea]. Disponible en: https://arxiv.org/abs/1506.02640.

4. Howard, A., Sandler M., Chu G., Chen L.-C., Chen, B., Tan, M., Wang W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., Adam, H. (2019). Searching for MobileNetV3. Accedido: 22 de octubre de 2023. [En línea]. https://doi.org/10.48550/arXiv.1905.02244.

5. Zhang, L., Zhao, C., Feng, Y., Li, D. (2023). Pests Identification of IP102 by YOLOv5 Embedded with the Novel Lightweight Module. Accedido: 23 de octubre de 2023. [En línea]. Disponible en: https://www.mdpi.com/2073-4395/13/6/1583.

6. Khalid, M., Sarfraz, M.S., Iqbal, U., Aftab, M.U., Niedbała, G., Rauf, H.T. (2023). Real-Time Plant Health Detection Using Deep Convolutional Neural Networks. Agriculture,13(2). Accedido: 24 de octubre de 2023. [En línea]. https://doi.org/10.3390/agriculture13020510.

7. Adnan, M. Mustafa, J. (2020). Plant Leaf Diseases Detection and Classification Using Image Processing and Deep Learning Techniques. *International Conference on Computer Science and Software Engineering*. Accedido: 24 de octubre de 2023. [En línea] https://ieeexplore.ieee.org/abstract/document/9142097.

8. Roboflow. (s.f). Accedido: 18 de septiembre de 2023. [En línea]. Disponible en: https://roboflow.com/.

9. GoogleColaboratory. Accedido: 18 de septiembre de 2023. [En línea]. Disponible en: https://colab.research.google.com/.

10. HWiNFO Diagnostic Software. Accedido: 18 de septiembre de 2023. [En línea]. Disponible en: https://www.hwinfo.com/download/.

SOBRE LOS AUTORES



Jesús García Palomera es Ingeniero en Mecatrónica con especialidad en Sistemas Embebidos. Actualmente es estudiante de la Maestría en Sistemas Computacionales en el Tecnológico Nacional de México campus Colima. Sus áreas de interés son la aplicación de la Inteligencia Artificial y sistemas de Visión por Computadora.



Jesús Alberto Verduzco Ramírez es Doctor en Informática por el Instituto Nacional Politécnico de Grenoble, Francia. Desde 1993 ha participado en labores administrativas y docentes del Instituto Tecnológico de Colima, habiendo dirigido tesis de posgrado y tesis de licenciatura. Ha publicado artículos de divulgación y difusión científica en revistas arbitradas y en congresos nacionales e internacionales. Ha participado en proyectos de investigación y en actividades colegiadas.



Luis Eduardo Morán López es maestro en ciencias de la computación, profesor por asignatura en la Facultad de ingeniería mecánica y eléctrica en la Universidad de Colima. Su línea de investigación es la visión artificial aplicada al reconocimiento gestual.



Juan García Virgen obtuvo el grado de Maestro en Ciencias en la Universidad de Colima (2000). Ha participado durante 20 años en labores administrativas y docentes en el Instituto Tecnológico de Colima, habiendo dirigido tesis de posgrado y tesis de licenciatura. Ha participado en proyectos de investigación y en actividades colegiadas. Sus áreas de interés en la investigación son las Tecnologías web y el Cómputo de Alto Rendimiento.



Imagen tomada de:
https://www.interempresas.net/Horticola/Articulos/383068-Una-nueva-aplicacion-detecta-de-forma-prematura-plagas-en-cultivos.html .

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

ARTÍCULO ACEPTADO

El Arte de los Origamis como Actuadores en la Robótica Suave

Juana-Mariel Dávila-Vilchis, Gustavo Levy Lara Morales, Adriana Herlinda Vilchis González, Melissa Días- Omonte, Juan Carlos Ávila Vilchis y Juan Manuel Jacinto-Villegas

Introducción

La palabra origami proviene del japonés ***Ori*** (doblez) y ***Kami*** (papel), cuyo origen se identifica como una actividad de entretenimiento. Este arte consiste en construir estructuras geométricas 3D de animales, flores u objetos con locomoción propia mediante diferentes dobleces secuenciales en papel, sin dañarlo [1]. Actualmente, esta técnica se ha aplicado en diversas áreas como Ingeniería, Arquitectura o Medicina. Por ejemplo, en el diseño de construcciones, barcos, transportadores espaciales o dispositivos médicos.

Otra aplicación de los origamis es en robótica suave donde se busca el desarrollo de dispositivos que estén dotados con movimientos prediseñados, múltiples grados de libertad, que sean capaces de emular movimientos naturales durante su accionamiento y que estén hechos de materiales ligeros [2]. En este contexto, un robot suave es accionado por medio de un *Origami Soft Actuator* (OSA) para ejecutar una tarea específica, que es controlada por un microcontrolador y una interfase, para ser empleado por un usuario. Los diseños poseen diferentes estructuras y morfologías, mayormente inspiradas en sistemas biológicos. Sus arquitecturas son variables dependiendo de un patrón geométrico, del número de capas a usar, de la técnica de doblez, del tipo de papel empleado y del espesor de este. Sus configuraciones incluyen modelos individuales, compuestos o modulares [2].

Las principales ventajas de los OSA son su versatilidad de formas, las cuales se caracterizan por tener estructuras compactas con la capacidad de ser reconfigurables a múltiples escalas. Además, al poderse doblar, se reducen el número de componentes y los costos gracias a la simplicidad de fabricación [3].

Los diseños OSA para robots médicos de rehabilitación o asistencia son variados, debido a la movilidad deseada, al número de grados de libertad (GDL), y a las tareas objetivo. Por ejemplo, el movimiento de flexo-extensión para miembros superiores se ha logrado por medio de los patrones de Zigzag y de Yoshimura. Para realizar en conjunto los movimientos de extensión, contracción y rotación se han empleado las estructuras Twisted Tower o Kresling.

Los origamis se pueden construir por dobleces a partir de un patrón 2D en papel o láminas de metal o plástico. Actualmente, la manufactura aditiva y la técnica de prototipado rápido utilizan poliuretano termoplástico (TPU), ácido poliláctico (PLA) o acrilonitrilo butadieno estireno (ABS) en la construcción de origamis. Por otra parte, los origamis pueden ser accionados por cables, por sistemas neumáticos de vacío, por materiales piezoeléctricos, Shape Memory Alloys (SMA) o Shape Memory Polymers (SMP) y su control está enfocado en regular la presión o las fuerzas para alcanzar el doblez deseado [4].

La literatura se ha enfocado en el diseño, modelado, simulación, fabricación y control de sistemas OSA [5], lo que motiva a presentar una metodología de caracterización de los origamis para el movimiento de flexo-extensión, con sus parámetros geométricos, proceso de impresión 3D, así como pruebas experimentales que validan su funcionamiento.

Los OSA son actuadores suaves, por lo que su uso en robótica suave los ubica en desafíos específicos de interacción segura con usuarios (pacientes o equipo médico, por ejemplo); para tareas de sujeción, regulación o seguimiento de trayectorias (flexo-extensión, por ejemplo) o de reconfiguración en tareas específicas como cirugía.

Origamis: Arte Japónes en Robótica Médica

En este documento se establecen guías de diseño que facilitan el desarrollo de futuros origamis aplicados a dispositivos médicos u otras áreas que busquen trabajar con actuadores suaves.

Diseño de los origamis

El primer paso para crear un Soft Origami, es identificar la naturaleza del movimiento a realizar para la correcta definición de su tipo, diseño y material, ya que pudiera ser frágil y romperse, o ser grueso y no poder doblarse. Además, es necesario determinar la fuerza requerida a aplicar, así como el tipo de actuación para lograrla. A continuación, se presenta una descripción detallada de 3 origamis neumáticos, 2 para flexo-extensión (Zigzag y Yoshimura) y 1 para rotación (Kresling), donde se incluye el diseño, fabricación y caracterización de sus

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

principales parámetros, como referente.

Los origamis neumáticos se contraen o expanden por medio de deflación o inflación, respectivamente. Requieren de una máquina de vacío o de un compresor para su accionamiento [6], pueden ser colocados dentro de contenedores plásticos sellados por calor o en estructuras 3D que contienen aire, formando n capas.

Origamis Zigzag: Se caracterizan por tener dobleces triangulares invertidos definidos por su longitud total L, su longitud del doblez l, y un ángulo interno (por ejemplo 60°) [7] (ver Figura 1). Es posible realizar movimientos de flexo-extensión dentro de bolsas plásticas selladas que funcionan al vacío y pueden soportar diferentes cargas.

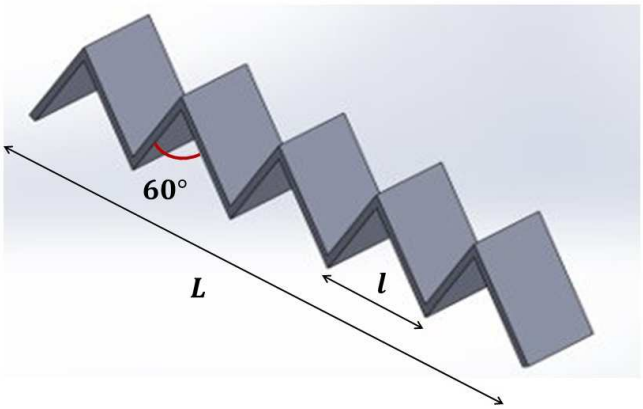


Figura 1. Modelo Zigzag (5 capas) en 3D.

Origamis Yoshimura: Su patrón es de una sola pieza. Se caracteriza por formar caras romboidales durante su doblez o impresión 3D. Está definido por un rectángulo de longitud L y ancho w, divido por 6 triángulos de 30°, cuyos bordes coinciden con pliegues en cresta o valle al ir doblándose (ver Figura 2). Puede ser accionado dentro de una bolsa plástica o contener aire en su interior para lograr la flexo-extensión [4].

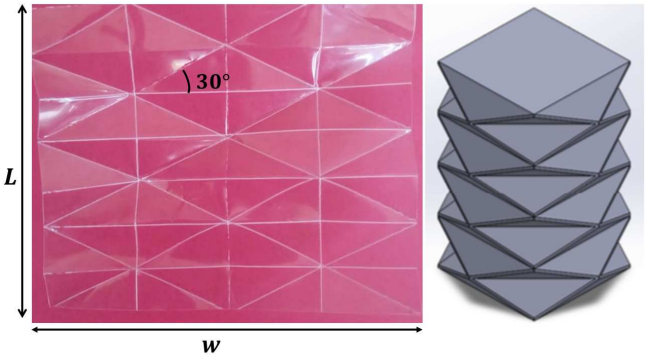


Figura 2. Patrón Yoshimura (4 capas) en acetato de vinil y modelo 3D.

Origamis Kresling: Tienen forma de espiral con 2 GDL: traslación y rotación simultánea. El modelo consta de dos polígonos en sus extremos, generalmente octagonales, conectados en sus vértices por un patrón de paralelogramos en una sola dirección formando n capas [4]. Lo caracterizan varios parámetros como longitudes, alturas, radios, ángulos de giro, número de caras y número de capas (ver Figura 3).

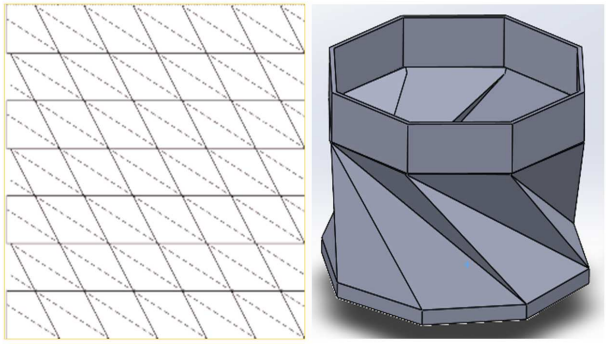


Figura 3. Patrón Kresling (1 capa) y modelo 3D.

Fabricación de los origamis

Se emplearon los métodos de Doblez y de Impresión 3D para la fabricación de los modelos anteriores. Para ambos se utilizaron materiales con diferentes propiedades. La Figura 4 muestra los productos obtenidos por ambos métodos y cambios en sus parámetros: diferentes tamaños, número de capas y materiales, tanto con doblez (1-3) como en impresión 3D (4-6).



Figura 4. Modelos 3D Yoshimura, Kresling y Zigzag en doblez (1-3) e impresión con TPU (4-6).

1. Doblez: Consiste en doblar el patrón de cada modelo para obtener una estructura 3D de n capas. Se hicieron pruebas con papel de espesores de 0.11 mm, 0.13 mm, 0.20 mm y 0.32 mm y con vinil de acetato de 0.34 mm. Se construyeron modelos de diferentes tamaños. Usando papel opalina de 0.20 mm se logró el mejor doblez y la mejor rigidez.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

2. Impresión 3D: A partir de un modelo CAD, se generó un archivo con extensión stl (Stereolithography) y otro con extensión gcode para impresión 3D. Para este trabajo se utilizó el software UltimakerCura® y la impresora Ender 3sl®, definiendo parámetros de velocidad, espesor y temperatura. El modelo Zigzag se imprimió con TPU-95A SAINSMART®y, posteriormente, se realizó el sellado de su bolsa y el ensamble de conectores y manguera. Para el modelo Yoshimura se utilizó TPU-85A NinjaTek®y para el modelo Kresling TPU-92A Creality®.

Caracterización de los origamis

Se hicieron pruebas experimentales de vacío para caracterizar el comportamiento de los origamis, donde se utilizaron un compresor TRUPER® de 58 psi, un eyector de vacío VN-10-H-T3PQ2-VQ2-RQ2 FESTO®y manguera de 10 mm de diámetro interior. La Tabla 1 muestra los parámetros de diseño de cada uno de los modelos Zigzag (Zz), Yoshimura (Y) y Kresling (K) y sus valores que incluyen: tipo (Ti), dimensiones (Di), dureza del material (Ma), espesor teórico (Es), porcentaje de relleno de impresión (Re), número de capas (Nc) y evaluación de su desempeño (D): funcional ($\surd$) o no funcional (x).

Tabla 1. Parámetros de modelos impresos.

Ti	Di [mm]	Ma	Es (mm)	Re %	Nc	D
Zz	150x30	TPU 95A	1	20	5	√
Y1	30x30x51	TPU 85A	1	20	4	√
Y2	30x30x36	TPU 85A	1	20	3	√
Y3	30x30x51	TPU 85A	0.5	20	4	√
Y4	50x50x208	TPU 92A	0.5	20	10	x
Y5	50x50x82	TPU 92A	0.8	20	4	√
Y6	25x25x97	TPU 92A	1	20	4	x
K1	40x54x54	TPU 92A	0.90	20	1	√
K2	24x24x60	TPU 92A	0.45	20	4	√
K3	54x54x120	TPU 85A	0.4	20	4	x

Resultados

Se caracterizaron experimentalmente los parámetros de diseño en cada uno de los origamis con diferentes configuraciones.

De estas pruebas se determinó que cada uno de los parámetros define la calidad de la pieza final para su correcto funcionamiento. Por ejemplo, para espesores delgados se requiere menor presión de succión; sin embargo, las piezas son frágiles y se rompen si la presión de succión excede un cierto valor o si se presentan áreas con rupturas relativamente de fácil generación (uniones).

El origami Zigzag fue el de diseño más simple y permitió levantar un candado de 250 g, tal como se observa en la Figura 5, desde su posición original hasta su contracción final. La capacidad de carga máxima dependerá del material utilizado, del calibre de la bolsa y de la presión de vacío.

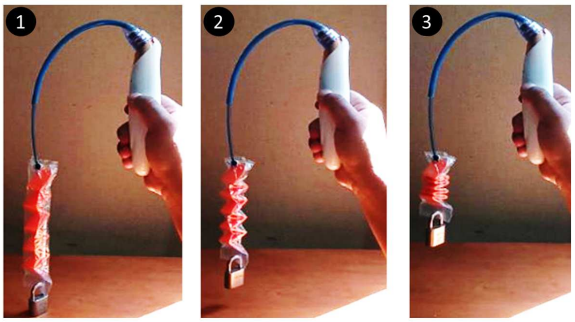


Figura 5. Origami Zigzag: 1) posición original, 2) extracción de vacío, 3) contracción total.

Los origamis Yoshimura y Kresling impresos en 3D también se pudieron elongar y contraer satisfactoriamente. La Figura 6 muestra el modelo Kresling en prueba de vacío: 1) origami en su estado inicial, 2) origami en contracción y 3) eyector de vacío. La regulación del aire se realizó manualmente para estas pruebas, por lo que, bajo un sistema de control automático, se esperarían resultados aún más satisfactorios. Este tipo de pruebas permite determinar rangos de operación de presiones en condiciones seguras.

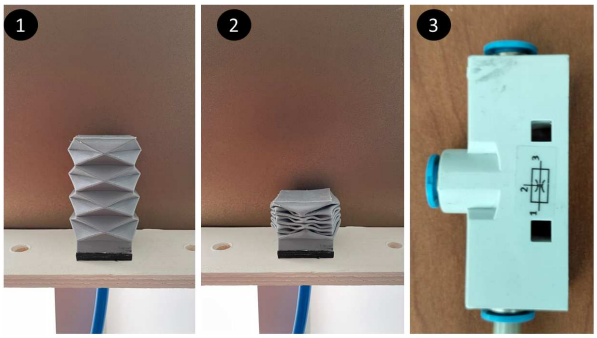


Figura 6. Sistema de pruebas de vacío.

Conclusiones

Este artículo reporta un panorama general de diseño, fabricación y caracterización de los origamis Zigzag, Yoshimura y Kresling para accionamiento de robots suaves. Estos actuadores pueden emplearse en tareas de flexo-extensión o rotación gracias a su capacidad para reconfigurarse. Sus parámetros de diseño pueden variar dependiendo del rango de movilidad y del tipo de movimiento deseados. Por ejemplo: largo, ancho, diámetro, número de capas, espesor y tipo de material deben ser definidos detalladamente.

Algunas de las métricas de interés para evaluar el desempeño de actuadores suaves basados en origamis son la capacidad de carga, el ciclo de vida, la rigidez y los retardos inherentes al proceso de deformación/recuperación de cada configuración, por ejemplo.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

Estos actuadores pueden generar movimientos de traslación, de rotación o la combinación de estos. Como aplicaciones médicas se identifican la rehabilitación física y ocupacional, la asistencia, o la dosificación de medicamentos. El proceso de fabricación de estos actuadores, su dependencia de un generador de vacío o de un compresor y la relativa fragilidad de los materiales utilizados son las principales limitantes para su uso.

Para la impresión 3D se recomienda usar filamento con dureza mayor a 85A y menor o igual a 92A debido a su facilidad de impresión con extrusor directo, ya que estos filamentos son más flexibles que el 95A. Al utilizar el TPU 95A de marca Creality®, se realizaron varias pruebas teniendo que disminuir el espesor de las paredes durante la impresión para lograr que los modelos tuvieran un grado de flexibilidad aceptable. Se sugiere imprimir una torre de temperatura para el filamento seleccionado antes de imprimir un modelo, para determinar el valor de temperatura que garantice la fusión del material entre capas.

El uso de origamis en sistemas robóticos complejos se ha considerado a través de sistemas modulares, donde la combinación de diferentes origamis generaría una variedad de GDL para la realización de tareas complejas (como rotaciones sucesivas de un bisturí, por ejemplo) por la interacción de dichos módulos. Una aplicación potencial de estos sistemas complejos modulares sería la cirugía mínimamente invasiva, donde un sistema robótico basado en origamis modulares se introdujera plegado al cuerpo humano y, una vez en su interior, se desplegase para realizar las tareas requeridas.✷

REFERENCIAS

1. Zhai, Z., Wang, Y., y Hanqing Jiang. (2018). Origami-inspired, on-demand deployable and collapsible mechanical matemetarials with tunable stiffness, 115(9), 1–6.
2. Zaghloul, A. y Bone, G. M. (2023). Origami-Inspired Soft Pneumatic Actuators: Generalization and Design Optimization. *Actuators*, 12(2), 72.
3. Fei, F., Leng, Y., Xian, S., et al. (2022). Design of an Origami Crawling Robot with Reconfigurable Sliding Feet. *Applied Sciences*, 12(5), 2520.
4. Son, H., Park, Y., et al (2022). 4D Multiscale Origami Soft Robots: A Review. Polymers, 14(19), 4235.
5. Rus, D. y Tolley, M. T. (2018). Design, fabrication and control of origami robots. *Nature Reviews Materials*, 3(6), 102-112.
6. Zhang, Z., Fan, W., Chen, G., et al. (2021). A 3D printable origami vacuum pneumatic artificial muscle with fast and powerful motion. En *Proc. 4th International Conference on Soft Robotics*, 551–554.
7. Li, S., Vogt, D. M., Rus, D. y Wood, R. J. (2017). Fluid-driven origami-inspired artificial muscles. *Proceedings of the National Academy of Sciences of the United States of America*, 114(50), 13132-13137.

SOBRE LOS AUTORES



Juana-Mariel Dávila-Vilchis es Doctora en Ciencias de la Ingeniería por la Universidad Autónoma del Estado de México (UAEM). Sus áreas de interés son la robótica suave y tecnología médica aplicada a tareas de rehabilitación y/o asistencia. Actualmente, es académica e investigadora de la Facultad de Ingeniería de la UAEM.



Gustavo Levy Lara Morales es alumno del 8° semestre de la Licenciatura de Ingeniería Mecánica, de la Facultad de Ingeniería de la UAEM. Sus intereses son el Diseño Mecánico en CAD, las técnicas de impresión 3D, así como el desarrollo de los origamis soft actuators.



Adriana Herlinda Vilchis González sus estudios de Doctorado en el Instituto Politécnico de Grenoble, Francia, están enfocados a Visión y Robótica. Actualmente, es profesor investigador en las Facultades de Ingeniería y de Medicina de la UAEM. Sus intereses incluyen robótica médica, robots suaves y sistemas mecatrónicos.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691



Melissa Días Omonte es alumna del 8° semestre de Ingeniería Generalista en la Escuela de Ingenieros de Centrale Marseille en Francia. Como parte de sus estudios realizó una estancia de investigación en la Facultad de Ingeniería de la UAEM, con el proyecto uso de origamis en la robótica suave.



Juan Carlos Ávila Vilchis obtuvo el grado de Doctor en el Instituto Nacional Politécnico de Grenoble, Francia. Actualmente es profesor investigador en las Facultades de Ingeniería y de Medicina de la UAEM. Sus intereses científicos incluyen modelado, análisis y control de sistemas no lineales y robótica médica.



Juan Manuel Jacinto Villegas Doctor en Tecnologías Digitales Emergentes en Robótica Perceptual de la Scuola Superiore Sant'Anna, Pisa, Italia. Sus áreas de intereses incluyen robótica, control háptico, interacción humano-robot, electrónica, mecatrónica y automatización.

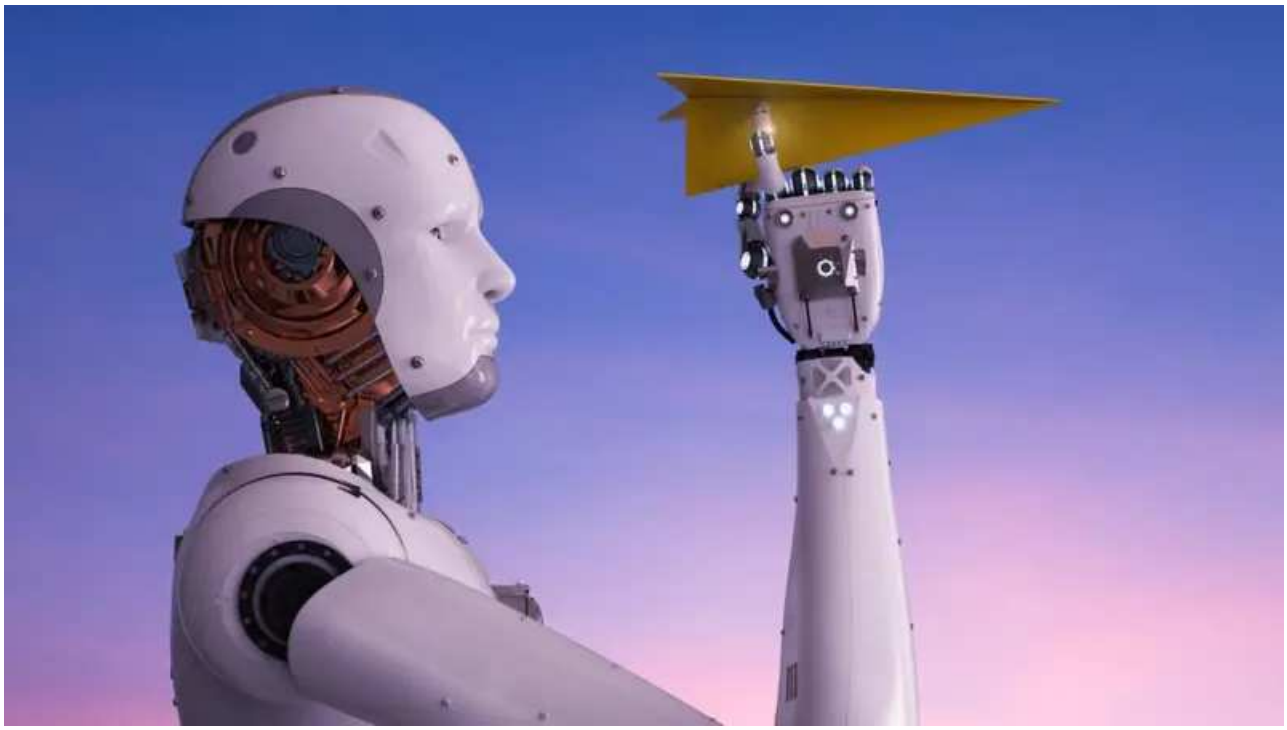


Imagen tomada de: https://www.bbc.com/mundo/noticias-50561369.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

ARTÍCULO ACEPTADO

Detección de ataques *phishing* mediante técnicas de inteligencia artificial

Lázaro Bustio-Martínez, Vitali Herrera-Semenets, Jorge Á. González-Ordiano y Miguel Á. Álvarez-Carmona

Resumen

El término *phishing* se refiere a un tipo de delito que utiliza canales de comunicación electrónicos para llevar a cabo ataques de ingeniería social con el objetivo de obtener información confidencial de los usuarios. Estos ataques suelen manipular a las víctimas generando sensaciones de urgencia, necesidad, vulnerabilidad, duda o curiosidad. La efectividad de estos ataques es significativa, ya que se dirigen al eslabón más vulnerable de la cadena de seguridad: las personas. Para contrarrestar estos ataques, es fundamental emplear métodos que puedan "aprender" de ataques previamente conocidos y aplicar estos aprendizajes a situaciones nuevas. En este contexto, la inteligencia artificial proporciona los métodos necesarios capaces de identificar ataques de *phishing* de día cero. Este artículo ofrece un recorrido histórico del surgimiento y evolución de los ataques de *phishing* y su mitigación desde la perspectiva de la inteligencia artificial.

Introducción

Con el auge de Internet, muchos aspectos de la sociedad evolucionaron hacia sus versiones en línea. Esta informatización también ha llevado a la evolución de los delitos hacia sus formas digitales, dando lugar al concepto de "ciberdelito". Aunque no existe una definición única para "ciberdelito", se acepta como tal aquellos actos ilegales dirigidos contra dispositivos electrónicos, redes y datos, así como acciones que utilizan internet y/o tecnología para atacar a personas. Uno de estos ciberdelitos, el *phishing*, ha cobrado especial relevancia debido a la pandemia de SARS-CoV-2 (COVID-19), destacándose por su peligrosidad y éxito. El *phishing* se caracteriza por comunicar mensajes de ingeniería social a las víctimas a través de canales de comunicación electrónicos, con el fin de persuadirlas para que realicen acciones en beneficio del atacante. Algunos ejemplos de información valiosa que busca un atacante, y que podrían usarse en una suplantación de identidad[1], son: i) detalles de la tarjeta de crédito (incluidos los datos de vencimiento y los códigos de seguridad de la tarjeta); ii) dirección postal y/o electrónica; iii) números de identificación, número de seguridad social y fecha de cumpleaños; iv) nombres de usuario y contraseñas de ciertos servicios; y v) datos de pasaporte; entre otros.

Para lograr su objetivo, es común manipular a las personas provocando sensaciones que disminuyan o incluso anulen los procesos psicológicos que conforman el sentido común y la suspicacia. Estas sensaciones pueden ser inducidas mediante el uso de los llamados "principios de persuasión", que son técnicas psicológicas utilizadas para influir en las decisiones de otras persona [1]. En este sentido, los atacantes se aprovechan de las técnicas y herramientas de la ingeniería social.

Los ataques de *phishing* se han convertido en una amenaza altamente efectiva y rentable porque se enfocan en atacar a las personas. Con el uso de nuevas tecnologías esta tendencia ha aumentado, y lo anterior se vio potenciado debido a la pandemia de COVID-19. En un estudio realizado por el *Antiphishing Working Group* (APWG) se concluyó que el número de ataques de *phishing* ha crecido de manera sostenida desde marzo de 2020, siendo el *phishing* dirigido a los servicios de correo electrónico la categoría más empleada, con el 31,4 % de todos los ataques [2].

Por todo lo anterior se hacen evidentes 2 hechos: 1) los métodos actuales de detección de ataques de *phishing* no son efectivos ni eficaces ante los retos que los ataques de *phishing* imponen en la actualidad; y 2) los medios sociales (tales como redes sociales, sistemas mensajería instantánea, correo electrónico, etc.) son los métodos más empleados por los atacantes para desarrollar campañas de *phishing* debido a la penetración social que estos medios tienen.

Además de lo anteriormente mencionado, es importante destacar que las campañas de *phishing* requieren un considerable tiempo de preparación en comparación con el tiempo necesario para su ejecución. Mientras que los ataques más genéricos pueden desarrollarse en un período relativamente corto, aquellos más específicos y elaborados, como el *spear phishing* dirigido hacia un objetivo particular, demandan un proceso de preparación mucho más extenso. Este tipo de ataques implica un estudio detallado del objetivo, incluyendo sus patrones de comportamiento, tendencias y historial, antes de poner en marcha la estrategia de ataque. Por lo tanto, su pre-

[1] En búsquedas recientes en la dark net realizada por los autores se determinó que una licencia de conducción mexicana se comercializa en 0.000077 BTC (bitcoins), lo que corresponde a 5 euros; mientras que un pasaporte mexicano tiene precios variables, comenzando desde los 0.0023 BTC, que son 150 euros.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

paración es significativamente más prolongada que la de los ataques más simples, pero también es probable que sean mucho más efectivos en términos de su impacto. La Figura 1 muestra el ciclo de vida de un ataque de *phishing*.

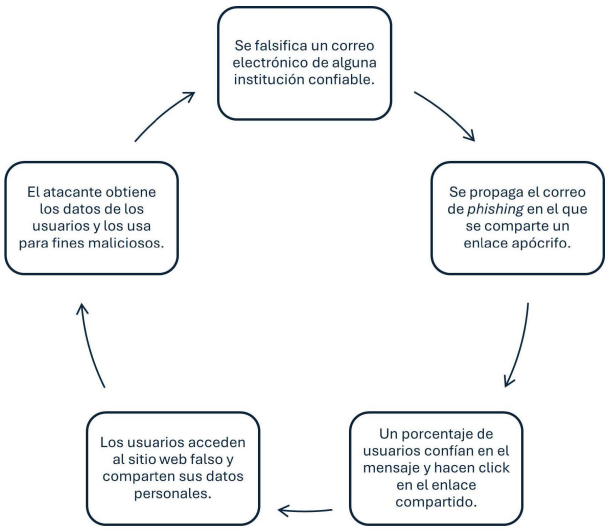


Figura 1. Ciclo de vida de un ataque de *phishing*.

El presente artículo recorre la historia de los ataques de *phishing*, desde su surgimiento hasta la actualidad, haciendo énfasis en cómo la inteligencia artificial está siendo utilizada para proponer nuevos métodos para la detección de estos ataques.

Definición y tipos de ataques de *phishing*: una visión integral

Phishing es un término derivado de la palabra inglesa *fishing* (o ir de pesca), donde el dígrafo *ph* hace referencia a *phone hacking*, una de las primeras formas de estafa. Tal como su nombre indica, los atacantes utilizaban llamadas telefónicas para llevar a cabo sus estafas. En la actualidad, *phishing* describe al fraude cibernético en el que los atacantes intentan obtener información confidencial (como contraseñas, números de tarjetas de crédito o detalles de cuentas bancarias) suplantando entidades legítimas a través de comunicaciones electrónicas fraudulentas. Este engaño se lleva a cabo mediante el uso de correos electrónicos, mensajes de texto, llamadas telefónicas u otros medios de comunicación, y se basa en la manipulación psicológica de las víctimas para que divulguen información sensible o descarguen *malware*. El objetivo principal de los ataques de *phishing* es inducir a las víctimas a realizar acciones que beneficien al atacante [3].

La ingeniería social es un área multifacética que abarca diversas disciplinas y técnicas para influir en el comportamiento humano. Si bien ha ganado notoriedad por su aplicación en tácticas abusivas para sorprender a personas incautas, es importante destacar que también tiene aspectos positivos y legítimos en contextos como el *marketing*, las ventas y la psicología aplicada.

En esencia, la ingeniería social implica el estudio y la manipulación de las interacciones humanas, aprovechando la confianza, la curiosidad, el miedo y otras emociones para lograr ciertos objetivos. Si bien su aplicación maliciosa ha sido ampliamente documentada en el contexto del *phishing* y otras formas de fraude cibernético, también se utiliza de manera ética para mejorar la comunicación, la persuasión y la interacción social en diversos campos.

En el contexto de los ataques de *phishing*, la ingeniería social desempeña un papel crucial, ya que los atacantes utilizan técnicas persuasivas para engañar a las víctimas y hacer que revelen información confidencial. Sin embargo, es importante reconocer que el estudio de la ingeniería Social va más allá de estos aspectos negativos, abarcando una amplia gama de estrategias para comprender y influir en el comportamiento humano de manera ética y efectiva.

Dentro del amplio espectro de los ataques de *phishing*, varios subtipos han surgido, cada uno con enfoques específicos y técnicas distintivas [3]:

1. ***Phishing tradicional***: Este es el método más común, en el cual los atacantes envían correos electrónicos masivos a una amplia audiencia, haciéndose pasar por instituciones financieras, empresas de renombre o servicios en línea populares. Estos correos electrónicos suelen contener enlaces maliciosos que redirigen a los usuarios a sitios web falsificados diseñados para robar información personal. Con este tipo de ataques lo que se busca es llegar a un grupo de víctimas más amplio, para maximizar la probabilidad de que algún desprevenido caiga en la trampa.

2. ***Spear Phishing***: A diferencia del *phishing* tradicional, el *spear phishing* implica un enfoque más dirigido y personalizado. Los atacantes investigan a fondo a sus objetivos y utilizan información específica sobre ellos para hacer que los mensajes sean más convincentes. Para este tipo de ataques, la información proveída por las personas de manera voluntaria en redes sociales suele ser crucial en la investigación y preparación del ataque. Esta técnica es especialmente efectiva contra empresas y organizaciones, donde los correos electrónicos pueden dirigirse a empleados de alto nivel con el fin de obtener acceso privilegiado o información confidencial.

3. ***Whaling***: Aunque similar al spear phishing, el *whaling* se centra en individuos de alto perfil, como

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

ejecutivos de empresas, políticos o celebridades. De ahí viene su nombre: *whaling*, en inglés relativo a ballena, o lo que es lo mismo: "peces gordos". En este tipo de ataque se busca obtener información confidencial o acceso a sistemas críticos mediante la suplantación de identidad de figuras importantes y el uso de técnicas de persuasión sofisticadas.

4. ***Vishing* (Voice Phishing)**: Este tipo de *phishing* implica el uso de las técnicas de ingeniería social mencionadas, pero a través de llamadas telefónicas para engañar a las víctimas. Los atacantes suelen hacerse pasar por representantes de servicios bancarios o instituciones gubernamentales, solicitando información confidencial o realizando acciones fraudulentas a través de manipulación verbal. El *vishing* puede llegar a ser muy efectivo, ya que implica una interacción directa y en tiempo real, lo que puede aumentar la presión sobre la víctima para que actúe rápidamente y comparta información confidencial. También, al ser una comunicación directa, suele ser muy adaptable en dependencia de la respuesta y reacción de las víctimas. En este sentido, los atacantes pueden explotar el empleo de emociones, como el miedo o la urgencia, durante una llamada telefónica, lo que puede resultar más convincente que un mensaje electrónico.

5. ***Smishing* (SMS Phishing)**: Aunque el uso de los SMS se ha visto declinado frente a las plataformas de comunicación instantáneas, como las redes sociales y chats, su uso ha cobrado una extrema relevancia como medio de verificación y autenticación de múltiples factores. El *smishing* utiliza mensajes de texto para engañar a las víctimas. Estos mensajes suelen contener enlaces maliciosos o números de teléfono falsificados, con el objetivo de robar información personal o inducir a las víctimas a descargar *malware* en sus dispositivos.

La Figura 2 muestra la relación entre las diferentes variantes de *phishing*.

Los ataques de *phishing* desde la historia

El surgimiento de los ataques de *phishing* se remonta a la década de 1990, cuando comenzaron a emerger los primeros intentos de engaño a través del correo electrónico. Desde su nacimiento en la década de 1990 hasta la fecha actual, los ataques de *phishing* han evolucionado considerablemente, aprovechando avances tecnológicos y estrategias cada vez más sofisticadas. A continuación, se explora el desarrollo y evolución histórica de los ataques de *phishing*, destacando los ejemplos significativos y sus consecuencias [4].

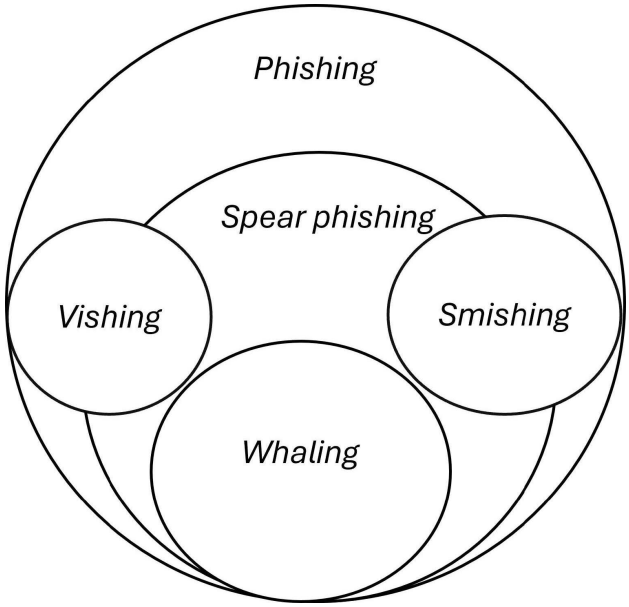


Figura 2. Diferentes variantes de *phishing* y su relación.

1990 - 2000: Los primeros pasos

Durante esta década, los ataques de *phishing* estaban en sus etapas iniciales y eran relativamente simples en comparación con las tácticas actuales. Uno de los primeros casos documentados fue el "AOL phishing scam" en 1996, donde los atacantes enviaban correos electrónicos falsos que parecían provenir del proveedor de servicios de internet AOL, solicitando a los usuarios que actualizaran su información de cuenta. Estos correos electrónicos llevaban a los usuarios a sitios web fraudulentos diseñados para robar credenciales de inicio de sesión. A medida que Internet se popularizaba, los ataques de *phishing* evolucionaron en complejidad y sofisticación, adoptando técnicas más avanzadas de ingeniería social y aprovechando vulnerabilidades en la seguridad de los sistemas informáticos y de red. Este aumento en la frecuencia y la gravedad de los ataques de *phishing* ha llevado a un interés creciente en el desarrollo de técnicas de detección automatizada para proteger a los usuarios y las organizaciones contra estas amenazas.

2003 - 2006: El surgimiento de la banca en línea

Con el crecimiento del uso de la banca en línea, los atacantes vieron una oportunidad lucrativa y comenzaron a dirigir sus ataques hacia instituciones financieras. En 2004, el "Rock Phish" se convirtió en uno de los kits de *phishing* más notorios, permitiendo a los estafadores crear fácilmente sitios web falsos que imitaban a bancos legítimos. Este kit se utilizó en numerosas campañas, incluida una que apuntaba al banco británico *Barclays* en 2006. Los usuarios engañados por estos sitios falsificados proporcionaban sus datos bancarios, lo que resultaba en pérdidas financieras significativas.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

2007 - 2010: El auge del *spear phishing*

El *phishing* evolucionó hacia formas más dirigidas y personalizadas con el surgimiento del *spear phishing*. El *spear phishing* es una forma altamente específica y dirigida de *phishing* que se diferencia de los ataques más genéricos al enfocarse en objetivos individuales o grupos selectos. En lugar de enviar correos electrónicos masivos a una amplia audiencia, los atacantes realizan una investigación detallada sobre sus víctimas para personalizar los mensajes de *phishing* de manera que parezcan legítimos y convincentes. Estos mensajes a menudo incluyen información personalizada, como nombres, cargos laborales o detalles sobre la empresa de la víctima, lo que aumenta la probabilidad de que la víctima caiga en la trampa. El *spear phishing* puede ser especialmente efectivo en entornos corporativos, donde los empleados pueden tener acceso a información confidencial o sistemas críticos, lo que lo convierte en una herramienta popular entre los ciberdelincuentes en busca de información sensible o acceso no autorizado.

En 2007, se informó que la campaña "Aurora" dirigida a empresas de tecnología y defensa fue una de las primeras instancias notables de *spear phishing* a gran escala. Los atacantes enviaron correos electrónicos aparentemente legítimos con archivos adjuntos infectados, aprovechando vulnerabilidades en el software para infiltrarse en las redes corporativas. Este incidente marcó una nueva era en la que los atacantes se dirigían a objetivos específicos con mensajes altamente personalizados.

2011 - 2015: Ataques de *phishing* a gran escala

Durante este período, se observaron varios ataques de *phishing* a gran escala que afectaron a millones de usuarios en todo el mundo. En 2011, la campaña "RSA SecurID" comprometió la seguridad de uno de los principales proveedores de autenticación de dos factores, resultando en el robo de información confidencial que podría comprometer la seguridad de numerosas organizaciones. En 2013, el ataque masivo a Adobe expuso las credenciales de 38 millones de usuarios, lo que llevó a un aumento significativo en el fraude y la identidad robada.

2016 - Presente: *Phishing* como Servicio y continua la sofisticación

En los últimos años los ataques de *phishing* han experimentado una evolución significativa, marcada por la emergencia de servicios especializados como el "Phishing as a Service" (PaaS), que ofrecen kits completos para llevar a cabo campañas de *phishing* de manera rápida y eficiente. Este fenómeno ha democratizado el acceso al *phishing*, permitiendo que incluso aquellos sin conocimientos técnicos profundos puedan perpetrar de manera exitosa estos ataques. Además, se ha observado una sofisticación continua en las técnicas de ingeniería social empleadas por los ciberdelincuentes, quienes aprovechan la urgencia y el miedo, como se vio durante la pandemia de COVID-19, para inducir a las víctimas a proporcionar información confidencial o hacer clic en enlaces maliciosos. Además, los ataques se han diversificado hacia nuevos canales de comunicación, incluyendo redes sociales y aplicaciones de mensajería instantánea, lo que amplía el alcance de los estafadores. Por último, para eludir las medidas de seguridad adicionales implementadas por organizaciones y usuarios, los ciberdelincuentes han comenzado a incluir elementos multifactoriales en sus ataques, como la suplantación de identidad combinada con solicitudes de códigos de autenticación de dos factores. Estos cambios destacan la importancia de mantenerse al día con las últimas tendencias en ciberseguridad y de implementar medidas de protección adecuadas para evitar ser víctima de estos ataques en constante evolución.

El *phishing* en los tiempos de COVID-19

Después del brote de COVID-19, los ataques de *phishing* han demostrado ser más prevalentes y sofisticados que nunca. Estos ataques se han adaptado a la pandemia, utilizando tácticas de ingeniería social que explotan la incertidumbre y el miedo relacionados con la crisis sanitaria global. Caracterizados por la suplantación de identidad de organismos de salud, empresas de entrega de alimentos y organizaciones gubernamentales, los correos electrónicos y mensajes de texto fraudulentos ofrecen información aparentemente urgente sobre la pandemia o sobre supuestas medidas de seguridad. Además, se ha observado un aumento en la diversificación de los canales de comunicación utilizados para el *phishing*, incluyendo plataformas de redes sociales y aplicaciones de mensajería instantánea, aprovechando la mayor actividad en línea durante los períodos de cuarentena y distanciamiento social. Estos ataques no solo tienen un impacto económico significativo para las organizaciones y usuarios, con pérdidas financieras estimadas en miles de millones de dólares, sino que también pueden tener consecuencias devastadoras para la seguridad y privacidad de los individuos y la integridad de las instituciones. En un mundo post-COVID-19, donde la digitalización se ha acelerado drásticamente, la ciberseguridad se ha convertido en una prioridad crítica para proteger tanto los datos personales como la infraestructura digital de ataques cada vez más persistentes y sofisticados.

La inteligencia artificial en la detección de ataques de *phishing*

En los primeros años del *phishing*, las soluciones de detección se basaban principalmente en listas negras de dominios y direcciones de correo electrónico conocidas por estar asociadas con actividades fraudulentas. Sin embargo, esta estrategia resultó ser insuficiente frente a la capa-

cidad de los atacantes para crear constantemente nuevas direcciones y dominios maliciosos.

Con el avance de la inteligencia artificial y el aprendizaje automático, surgieron nuevas técnicas para detectar ataques de *phishing* de manera más efectiva. En la década de 2000, los investigadores comenzaron a explorar el potencial del aprendizaje automático para identificar patrones en los correos electrónicos de *phishing* y distinguirlos de los correos legítimos [5].

Un hito importante en esta evolución fue el desarrollo de algoritmos de aprendizaje automático capaces de extraer características relevantes de los correos electrónicos, como la presencia de enlaces sospechosos, el uso de lenguaje persuasivo y la solicitud de información confidencial. Estos algoritmos se entrenaron utilizando conjuntos de datos de *phishing* etiquetados manualmente, lo que permitió mejorar su precisión y capacidad de detección [6]. Un ejemplo destacado de este enfoque es el uso de modelos de lenguaje natural en la detección de *phishing*. Al analizar el contenido y la estructura de los correos electrónicos, estos modelos pueden identificar indicadores clave de *phishing*, como la urgencia en el mensaje, la solicitud de información personal y la imitación de marcas reconocidas [7].

Otro avance significativo fue la aplicación de redes neuronales convolucionales y redes neuronales recurrentes para analizar el contenido visual y temporal de los correos electrónicos, respectivamente. Estas técnicas permitieron detectar elementos visuales como logotipos falsificados y analizar la secuencia de eventos en los correos electrónicos, lo que mejoró la detección de ataques sofisticados [8].

En la última década, el desarrollo de modelos de inteligencia artificial más avanzados, como las redes neuronales profundas y los modelos de lenguaje pre-entrenados, ha llevado la detección de *phishing* a un nivel superior de precisión y escalabilidad. Estos modelos pueden aprender de manera autónoma patrones complejos en grandes volúmenes de datos, lo que les permite adaptarse rápidamente a las tácticas cambiantes de los atacantes.

Sin embargo, es importante destacar que, además de ser utilizados para detectar intentos de *phishing*, los sistemas de inteligencia artificial también pueden ser empleados para crear ataques de *phishing* más sofisticados y convincentes. Los avances en inteligencia artificial han permitido a los atacantes automatizar la generación de correos electrónicos y sitios web falsificados, utilizando técnicas como la generación de texto natural y el diseño de interfaces de usuario convincentes. Esta dualidad en el uso de la inteligencia artificial presenta un desafío adicional en la lucha contra el *phishing*, ya que los mismos avances tecnológicos que mejoran la capacidad de detección también pueden ser aprovechados por los perpetradores para mejorar la efectividad de sus ataques.

Además de lo anteriormente expresado, es crucial reconocer que aunque las técnicas modernas de aprendizaje automático han demostrado ser efectivas en la detección de ataques de *phishing*, su funcionamiento como cajas negras plantea desafíos en términos de explicabilidad. La falta de transparencia en la toma de decisiones de estos modelos dificulta la comprensión de cómo se detectan los ataques, lo que puede limitar su utilidad para concientizar y educar a los usuarios. Por ello no se deben abandonar otros enfoques más clásicos que ofrecen soluciones directas y fáciles de entender, como los clasificadores basados en árboles de decisión o reglas. Al visualizar patrones de *phishing* representados como reglas, los usuarios pueden comprender mejor cómo funcionan los ataques. Por ejemplo, una regla podría señalar correos electrónicos que contienen enlaces a dominios falsos, solicitan información confidencial o presentan combinación de frases claves junto a determinados principios de persuasión. Estos patrones concretos ayudan a los usuarios a reconocer señales de advertencia.

Por otro lado, es fundamental diferenciar entre los sistemas de detección de *phishing*, cuyo objetivo principal es identificar y bloquear mensajes o correos electrónicos sospechosos, y los sistemas educativos destinados a empoderar a los usuarios para que reconozcan y eviten los ataques por sí mismos. Mientras que los primeros se centran en la protección activa de los usuarios al bloquear el acceso a contenido malicioso, los segundos tienen como objetivo principal aumentar la conciencia y la capacidad de respuesta de los usuarios ante posibles amenazas.

Sin embargo, a pesar de estos avances, la detección de *phishing* sigue siendo un desafío en constante evolución. Los atacantes continúan desarrollando nuevas técnicas para eludir los sistemas de detección, como el uso de técnicas de evasión y la personalización de los correos electrónicos para hacerlos más convincentes.

Conclusiones

La detección temprana de ataques de *phishing* resulta crucial para proteger a usuarios y organizaciones. La evolución ha sido un aspecto que a caracterizado la forma de realizar estos ataques en los últimos años. A su vez, los atacantes siguen innovando y adaptándose a los nuevos contextos y escenarios. Por lo tanto, la investigación y el desarrollo continuo en el campo de la detección de *phishing* son esenciales para no quedar resagados. La inteligencia artificial ha demostrado ser una herramienta efectiva para identificar patrones representativos de *phishing* en correos electrónicos y otros medios. Aunque estos modelos avanzados ofrecen buenos resultados, no se deben descartar enfoques más clásicos y comprensibles, como los clasificadores basados en reglas. Estos últimos, además de aportar una solución al problema, también pueden coadyuvar a la educación de los usuarios sobre las señales de advertencia de *phishing*. La inteligencia artificial y el aprendizaje automático pueden ayudar, pero

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

la conciencia humana sigue siendo crucial para fortalecer las defensas de los sistemas.✷

REFERENCIAS

1. Cialdini, R. B. (2007). *Influence: The psychology of persuasion*. Collins New York.
2. Anti-Phishing Working Group (APWG). (2024). *Phishing Activity Trends Report, 4th Quarter 2023*. Recuperado de https://www.apwg.org
3. Basit, A., Zafar, M., Liu, X., Javed, A. R., Jalil, Z., & Kifayat, K. (2021). *A comprehensive survey of AI-enabled phishing attacks detection techniques*. Telecommunication Systems, 76(1), 139-154. Recuperado de https://doi.org/10.1007/s11235-020-00733-2
4. Tandale, K. D., & Pawar, S. N. (2020). *Different Types of Phishing Attacks and Detection Techniques: A Review*. En 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC) (pp. 295-299). Recuperado de https://doi.org/10.1109/ICSIDEMPC49020.2020.9299624
5. Karim, A., Azam, S., Shanmugam, B., Kannoorpatti, K., & Alazab, M. (2019). *A Comprehensive Survey for Intelligent Spam Email Detection*. IEEE Access, 7, 168261-168295. Recuperado de https://doi.org/10.1109/ACCESS.2019.2954791
6. Bustio-Martínez, L., Álvarez-Carmona, M. A., Herrera-Semenets, V., Feregrino-Uribe, C., & Cumplido, R. (2022). *A lightweight data representation for phishing URLs detection in IoT environments*. Information Sciences, 603, 42-59. Recuperado de https://doi.org/10.1016/j.ins.2022.04.059
7. Misra, K., & Rayz, J. T. (2022). *LMs go Phishing: Adapting Pre-trained Language Models to Detect Phishing Emails*. En 2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT) (pp. 135-142). Recuperado de https://doi.org/10.1109/WI-IAT55865.2022.00028
8. Yerima, S. Y., & Alzaylaee, M. K. (2020). *High Accuracy Phishing Detection Based on Convolutional Neural Networks*. En 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS) (pp. 1-6). Recuperado de https://doi.org/10.1109/ICCAIS48893.2020.9096869

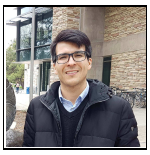
SOBRE LOS AUTORES



Lázaro Bustio-Martínez El Dr. Lázaro Bustio-Martínez obtuvo el grado de Ingeniero en Informática en el año 2006 por el Instituto Superior Politécnico "José Antonio Echeverría", en La Habana, Cuba. En el año 2010 obtuvo el título de Máster en Sistemas Computacionales por el Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE) en Puebla, donde también concluyó su doctorado en 2017. Actualmente, es Investigador Académico de tiempo completo en el Departamento de Estudios de Ingeniería para la Innovación en la Universidad Iberoamericana y miembro nivel 1 del Sistema Nacional de Investigadores. Sus líneas de investigación se orientan a la aplicación de técnicas de Inteligencia Artificial a la Ciberseguridad entre otras.



Vitali Herrera-Semenets El Dr. Vitali Herrera-Semenets obtuvo el grado de Ingeniero en Ciencias Informáticas en el año 2009 por la Universidad de Ciencias Informáticas, en La Habana, Cuba; donde también concluyó su doctorado en 2019. Actualmente, dirige el departamento de Minería de Datos del Centro de Aplicaciones de Tecnologías de Avanzada (CENATAV) en La Habana, Cuba. Además es miembro de la junta directiva de la Asociación Cubana de Reconocimiento de Patrones (ACRP) y miembro de la Asociación Internacional de Reconocimiento de Patrones (IAPR).



Jorge Á. González-Ordiano El Dr. Jorge Á. González-Ordiano es Ingeniero Mecánico egresado del *Karlsruhe Institute of Technology* donde también completó una maestría en Ingeniería Mecánica y un doctorado en Ingeniería. Participó en el proyecto *Energy System 2050* de la *Helmholtz Association*, desarrollando modelos de pronóstico de energía renovable. Como investigador posdoctoral en la *Colorado State University*, trabajó en modelos para combatir el comercio ilícito. Desde 2020, es académico en la Universidad Iberoamericana, y desde 2023, director del Instituto de Investigación Aplicada y Tecnología. Miembro nivel 1 del Sistema Nacional de Investigadores, su investigación se enfoca en el uso de ciencia de datos para resolver problemas diversos, desde transición energética hasta procesamiento de imágenes digitales.



Miguel Á. Álvarez-Carmona El Dr. Álvarez-Carmona estudió Ciencias Computacionales en la Benemérita Universidad Autónoma de Puebla donde se tituló con los máximos honores en 2013. Obtuvo los grados de maestro y doctor en Ciencia Computacionales en el Instituto Nacional de Astrofísica, Óptica y Electrónica en 2014 y 2019 respectivamente; donde se especializó en inteligencia artificial y procesamiento de lenguaje natural. Desde 2019 se desempeña como Investigador por México. Actualmente labora en el CIMAT Monterrey, haciendo investigación en procesamiento de lenguaje natural, aprendizaje de máquina y similitud de semánticas; aplicado en su mayoría a las ciencias sociales. El Dr. Álvarez pertenece al Sistema Nacional de Investigadores en el nivel 1 y a diversos grupos de investigación.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

ARTÍCULO ACEPTADO

Develando la detección de toxicidad en las plataformas digitales: El papel de las redes neuronales artificiales

Allan Balam Rueda-Gutíerrez, Alejandro Rodríguez-Molina, Oscar Armando Gómez-Vargas y Eric Hernández-Castillo

Introducción

En el contexto digital reciente, las personas pueden relacionarse a través de Internet y de plataformas en línea como los blogs y las redes sociales. Esto les permite expresar y compartir sus ideas, opiniones y pensamientos. Sin embargo, un problema con estos espacios digitales de expresión, es el incremento alarmante de mensajes y comentarios tóxicos y dañinos. Este tipo de prácticas ha provocado discursos de odio, acoso y desinformación, entre otras consecuencias negativas, por lo que su detección oportuna es necesaria para prevenir y atender los problemas que genera. La detección de textos tóxicos en plataformas digitales es una tarea compleja, por lo que resulta de interés develar los conceptos y desafíos alrededor de esta, así como los métodos tradicionales y los nuevos enfoques basados en inteligencia artificial que se utilizan para resolver los problemas asociados. Con esto se puede mejorar la comprensión del problema y de sus alternativas de solución, a fin de fomentar su uso para promover la seguridad, el respeto y las interacciones sanas en los espacios digitales.

El presente trabajo discute los enfoques utilizados para la detección de textos tóxicos, desde los métodos tradicionales hasta los basados en redes neuronales artificiales, que se utilizan ampliamente en la actualidad. Además, describe las etapas necesarias para desarrollar una red neuronal capaz de detectar toxicidad, comenzando con la adquisición de textos y finalizando con la evaluación de la eficacia del método.

La Toxicidad

La toxicidad de comentarios en las plataformas digitales se refiere a la medida en que su contenido puede causar daños a usuarios específicos o a comunidades. Este tipo de contenido perjudicial puede incluir amenazas, insultos, sexismo, opiniones extremistas, racistas y despectivas, etc. Algunos tipos de comentarios tóxicos pueden incluir lenguaje de odio acompañado de información privada de usuarios para causarles daño. Otros tipos pueden promover conductas de acoso criminal y violencia. Existen incluso algunos comentarios que causan daños de manera sutil, es decir, que el daño está implícito, pero resulta difícil de comprender o identificar [1].

Identificar el grado de toxicidad en los comentarios vertidos en medios digitales, implica detectar y clasificar de manera automática los textos que contienen características ofensivas y comprender las diferentes maneras en la que se pueden manifestar. Algunos estudios han planteado diferentes enfoques para resolver el problema de identificación de toxicidad que van desde el uso de reglas lingüísticas predefinidas, hasta técnicas avanzadas que utilizan modelos de inteligencia artificial, los cuales, son programas computacionales que buscan imitar la forma en que un humano experto puede detectar contenido tóxico. No obstante, dicho problema sigue siendo un desafío debido a la diversidad cultural y lingüística que existe en la comunidad digital. Por esta razón, es importante implementar estrategias efectivas que puedan hacer frente al problema y disminuir los niveles de toxicidad en los miles de mensajes generados a cada minuto en estas plataformas. Con esto, se pueden promover y garantizar entornos más seguros, respetuosos y de sana convivencia [2].

El enfoque tradicional para la detección de toxicidad en plataformas digitales se centra principalmente en el uso de reglas predefinidas que identifican palabras clave que se consideran ofensivas dentro de un texto. Al conjunto de estas palabras se le denomina lenguaje vulgar. Entonces, para identificar comentarios tóxicos, el método tradicional emplea diccionarios, reglas gramaticales y patrones que comúnmente se utilizan para ofender, como se ilustra en la Figura 1. Sin embargo, este tipo de enfoque se enfrenta a varios desafíos relacionados con la precisión, la evolución del lenguaje y la dificultad para detectar el uso de jergas [3].

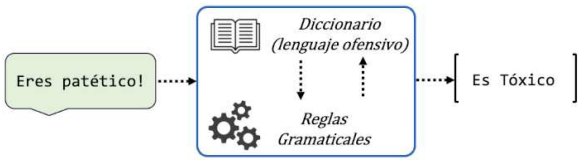


Figura 1. Método tradicional basado en lenguaje vulgar para detectar textos tóxicos.

Otro problema al que se enfrenta el método tradicional, es el de identificar el contenido basado en términos léxicos aislados [4], es decir, frases compuestas por palabras que aparentemente no son tóxicas, pero que jun-

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

tas pueden tener un significado distinto. Por ejemplo, la frase "te vistes como abuelito" muestra un contexto de desprecio cultural y de discriminación, pero no contiene palabras ofensivas, por lo que difícilmente sería detectada como tóxica por el método tradicional, como se muestra en la Figura 2. Otras frases relacionadas con el humor, las burlas y la ironía, son también difíciles de identificar como tóxicas usando este método [5].

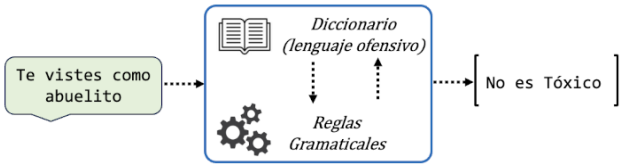


Figura 2. El problema de léxicos asilados en el enfoque tradicional.

Debido las limitaciones que posee el método tradicional y aunado a la interacción activa y creciente entre los usuarios de plataformas digitales, en especial, los de redes sociales, es evidente que se necesitan alternativas más sofisticadas para realizar una mejor detección de los niveles de toxicidad en textos. Una opción prometedora está basada en modelos de inteligencia artificial.

Métodos basados en Inteligencia Artificial

Los métodos que emplean modelos de inteligencia artificial orientados al procesamiento de lenguaje natural, es decir, que se centran en entender la manera en que se comunican los humanos, se consideran las alternativas más precisas y avanzadas para realizar la identificación de los niveles de toxicidad en textos y comentarios vertidos en las plataformas digitales. Entre los modelos que se utilizan con este fin resaltan las máquinas de soporte vectorial, el clasificador bayesiano ingenuo, la regresión logística, los árboles de decisión y los bosques aleatorios, el algoritmo de los K vecinos más cercanos, los modelos basados en memoria, los modelos de ensamblaje, y los transformadores. No obstante, las Redes Neuronales Artificiales (RNA) son una alternativa sobresaliente para llevar a cabo esta difícil tarea. Las RNA son capaces de reconocer y aprender patrones complejos en textos escritos por personas, y a clasificarlos con base en la semántica del contexto. Asimismo, estos modelos han logrado realizar detecciones y clasificaciones más precisas de toxicidad en textos debido a que pueden aprender de forma constante a pesar de la evolución del lenguaje y las nuevas formas de expresión de los usuarios de Internet [6].

Las RNA son modelos inspirados en la estructura biológica del cerebro humano, en donde se encuentran millones de células llamadas neuronas que se interconectan y comunican entre sí para responder a situaciones complejas y diversas con base en los estímulos que reciben como entrada. De forma simplificada, una neurona se ve como en la Figura 3 y consta del soma o cuerpo neuronal, que realiza el procesamiento de los estímulos; las dendritas, que son las vías de entrada de esos estímulos; y el axón, que constituye la vía de salida de la respuesta. En el ámbito computacional, una sola neurona puede ser abstraída y representada mediante un perceptrón como el que se observa en la Figura 4 [7]. Similar a la neurona biológica, el perceptrón procesa múltiples entradas x_i, cada una de ellas más o menos importante que el resto. Para indicar lo anterior, cada entrada x_i se multiplica por un valor numérico w_i llamado peso. Estos estímulos de entrada determinan si el perceptrón debe activarse para enviar una respuesta. Para esto, primero las entradas ponderadas $w_i x_i$ se suman y al resultado se le añade el término de sesgo b, que sirve como umbral para la activación del perceptrón. Luego, el valor obtenido en la etapa anterior se evalúa con una función de activación que finalmente decide la respuesta que dará la neurona artificial.

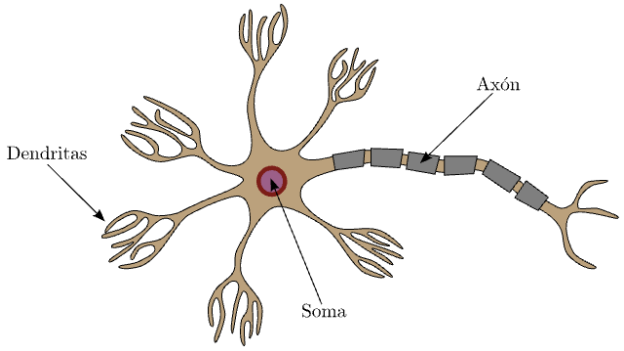


Figura 3. Neurona biológica del cerebro humano.

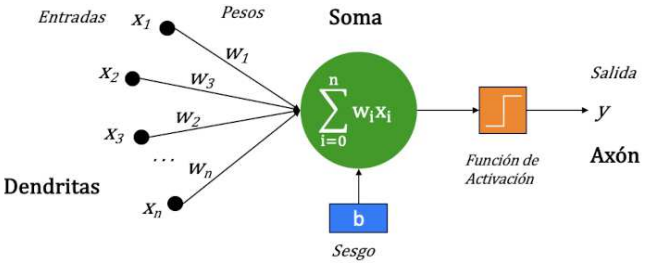


Figura 4. El perceptrón como la representación computacional de una neurona.

Redes Neuronales Artificiales

Las redes neuronales artificiales se componen de múltiples neuronas/perceptrones organizados en distintas capas, como se aprecia en la Figura 5. En la disposición de esa figura, las neuronas de la capa de entrada se encargan

de recibir los estímulos del exterior. Esta primera capa comunica la información de entrada a las neuronas de la siguiente capa conocida como oculta y, a su vez, esta última transfiere sus resultados a la siguiente. El proceso se repite hasta la última capa oculta. Cada neurona de las capas ocultas tiene como entradas los valores de salida de la capa inmediata anterior, realiza el proceso de ponderación usando sus propios valores de peso, agrega su propio término de sesgo y calcula un resultado utilizando su propia función de activación. Finalmente, en la capa de salida, se encuentran las neuronas que calculan el resultado final de la RNA. Para esto, las neuronas de esta capa toman la información de la última capa oculta y la procesan con sus propios pesos, términos de sesgo y funciones de activación [8].

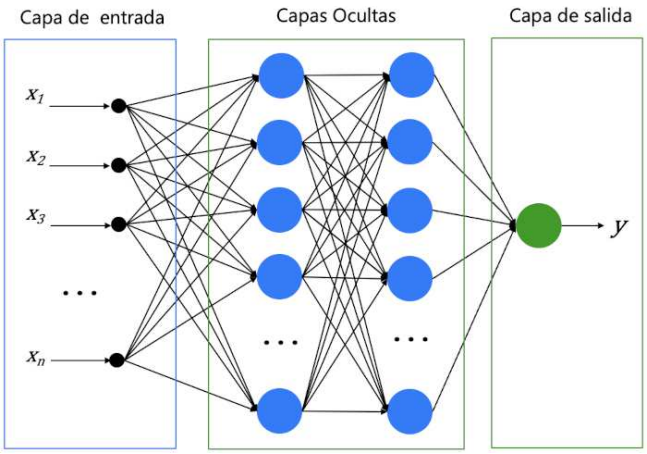


Figura 5. Red Neuronal Artificial.

Para que una red neuronal como la de la Figura 5 aprenda a reconocer patrones en datos, como los que están contenidos en comentarios de texto, es necesario llevar a cabo un proceso de entrenamiento. El entrenamiento consiste en ajustar los valores de los parámetros en la RNA. Éstos comúnmente incluyen a las ponderaciones de las entradas y al término de sesgo que posee cada neurona en la red (ver Figura 4). Durante el entrenamiento, se le proporciona a la RNA varios ejemplos reales de entradas (por ejemplo, oraciones de lenguaje natural) y salidas (por ejemplo, un valor que indique si las oraciones tienen contenido tóxico o no). Con base en esa información, los parámetros de la RNA se ajustan para tratar de reducir las diferencias entre las salidas que predice y las reales al recibir la entrada real de cada ejemplo.

Debido a su evidente flexibilidad y elevado poder computacional al agrupar distintas unidades de procesamiento como el perceptrón, las RNA son utilizadas para detectar patrones muy complejos en los datos, inclusive aquellos que tienen relación con contenidos tóxicos en textos.

En la actualidad, existen diferentes arquitecturas de RNA, es decir, diferentes formas de acomodar las neuronas, capas y conexiones en una red, que son capaces de identificar de mejor manera patrones de toxicidad en textos. Las arquitecturas más destacadas son las redes neuronales convolucionales, las redes neuronales recurrentes, las redes neuronales LSTM (Long Short-Term Memory, por sus siglas en inglés) y el modelo BERT de redes neuronales propuesto por Google y utilizado en el campo del procesamiento del lenguaje natural [9].

Para realizar la detección de toxicidad utilizando RNA, la información contenida en oraciones de lenguaje natural es acondicionada para incluirse como entrada de una red entrenada. Entonces, la red puede detectar patrones y clasificar los diferentes elementos de toxicidad que identifique. El proceso se describe en las siguientes fases:

1. Obtención de datos. Al momento de adquirir los datos de toxicidad con los que se entrenará la RNA, se deben considerar factores importantes como las fuentes de dicha información, las cuales deben ser variadas; la manera en que se recopilan y representan los datos, procurando siempre un formato uniforme y universal; y la cantidad y calidad de los mismos, tratando siempre de que la información sea diversa, es decir, que incluya suficientes textos tóxicos y no tóxicos representativos. Una forma conveniente para concentrar los datos de toxicidad es mediante un formato tabular simple y de fácil representación como el formato de valores separados por comas (CSV, por sus siglas en inglés). Dicho formato puede utilizarse fácilmente por la mayoría de las herramientas de inteligencia artificial actuales. En el contexto de los datos tóxicos, la información en formato tabular tiene un aspecto similar al de la Figura 6.

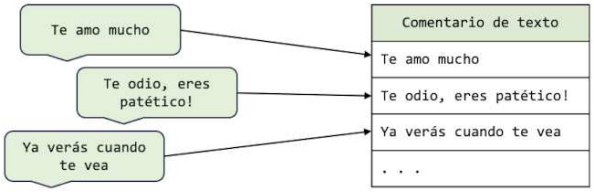


Figura 6. Datos de toxicidad en formato tabular.

2. El preprocesamiento de los datos. Esta fase se refiere al tratamiento y transformación de los datos en un formato óptimo y comprensible para que la RNA pueda aprender de ellos. En el contexto de detección de toxicidad, la RNA aprende a clasificar mensajes. La clasificación de datos es un problema fundamental en inteligencia artificial que consiste en predecir la categoría, clase o etiqueta a la que pertenece un dato nuevo o desconocido a partir de la información conocida de un conjunto de datos correctamente etiquetados (el etiquetado se realiza usualmente con la ayuda de un humano experto).

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

Detectar la toxicidad en comentarios ayuda a mantener una convivencia sana en comunidades digitales.

En la actualidad, existen conjuntos de datos efectivos que contienen mensajes de toxicidad extraídos de Internet y se han etiquetado para indicar el grado o nivel de toxicidad que contienen con la finalidad de construir modelos de inteligencia artificial. Las muestras que se incluyen en esos conjuntos oscilan entre 500 y 500,000. Las etiquetas que se colocan en cada mensaje son binarias, es decir, tienen valores de cero y uno que indican si un tipo de contenido tóxico está presente o no. De esta forma, se puede indicar claramente con uno o varios unos si un texto es considerado tóxico y en qué medida, es decir, el texto puede ser identificado como tóxico, tóxico severo, obsceno, amenazante, insultante o de odio. Si un texto está etiquetado solo con ceros, quiere decir que no presenta contenido tóxico. La Figura 7 ejemplifica la manera en que podrían etiquetarse los textos de la Figura 6.

Comentario de texto	Tóxico	Tóxico Severo	Obsceno	Amenaza	Insulto	Odio
Te amo mucho	0	0	0	0	0	0
Te odio, eres patético!	1	0	0	0	1	1
Ya verás cuando te vea	1	0	0	1	0	1
. . .						

Figura 7. Etiquetado de mensajes.

3. Tokenización. A diferencia de los mecanismos cognitivos que permiten a los humanos entender oraciones complejas, los modelos computacionales como las RNA solo son capaces de entender valores numéricos. Por ello, la detección de mensajes tóxicos requiere de un proceso adicional para transformar textos de cualquier tamaño en pequeñas piezas de valores numéricos que se pueden interpretar con una computadora. La tokenización es un término que se refiere a la tarea de preprocesamiento en la que se toma cada mensaje y se convierte en una secuencia de números que se conocen como tokens. Un token es una representación numérica de cada palabra contenida en un mensaje, como se ilustra en la Figura 8. Es importante realizar este proceso debido a que una RNA no puede trabajar directamente con texto o palabras, pero sí puede hacerlo con números.

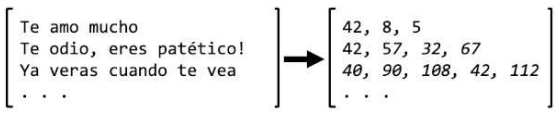


Figura 8. Tokenización de mensajes.

Después del proceso de tokenización se realiza la tarea de incrustación. Esta consiste en asignar a cada token una serie de valores reales que describen su significado semántico y contextual. Inicialmente, estos números se eligen al azar dentro de un rango específico, pero se ajustan durante el entrenamiento para reflejar mejor el significado y el contexto de las palabras. En este sentido, los tokens similares entre sí tienen representaciones numéricas cercanas. Las incrustaciones de números se usan como entrada para la red neuronal.

En la Figura 9, se muestra un ejemplo de este proceso con una serie de 4 números, aunque en la práctica, este tamaño puede variar desde 50 hasta 1024 números.

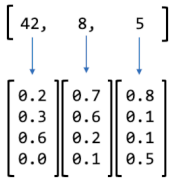


Figura 9. Tarea de incrustación.

4. Construcción del modelo de red neuronal. Este punto se refiere a la selección de una arquitectura de RNA que sea capaz de trabajar con secuencias de entradas, es decir, con incrustaciones. Aunque existe en la actualidad una gran variedad de esas arquitecturas, una de las más útiles para realizar la identificación y clasificación de los niveles de toxicidad en textos es la LSTM [6] aunque también se han propuesto arquitecturas híbridas que tienen una alta precisión al combinar LSTM con redes neuronales convolucionales [10]. Este tipo de RNA ha mostrado una elevada precisión al realizar la clasificación de mensajes. Como se puede observar en la Figura 10, una RNA de este tipo recibe como entrada de datos las incrustaciones que representan secuencias de tokens derivados de las palabras de un texto. Como toda RNA, cuenta con varias capas ocultas con las neuronas necesarias para realizar la predicción de una o varias clases. Las primeras capas de la red neuronal tienen la capacidad de recordar información de las incrustaciones anteriores para entender mejor el significado del texto. Luego, las capas intermedias se encargan de procesar esta información y sus relaciones. Finalmente, las capas de salida toman los datos procesados y los convierten en un resultado final.

El resultado es un conjunto de números que indican si un comentario es tóxico o no, basándose en categorías predefinidas. En la Figura 10, se muestran 6 salidas de

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

la red que corresponden a las 6 categorías de la tabla de la Figura 7. La primera salida indica si un texto es "Tóxico" y las demás categorizan el texto en otras etiquetas relacionadas. Cada salida funciona como un interruptor que indica "Sí" (1) o "No" (0) para cada categoría.

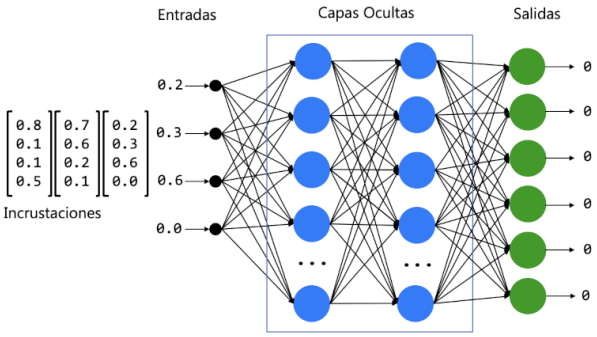


Figura 10. RNA utilizada para manejar incrustaciones.

5. Entrenamiento del modelo. El proceso de entrenamiento implica utilizar un conjunto de datos con múltiples ejemplos de entradas y salidas, es decir, con mensajes tóxicos y no tóxicos correctamente etiquetados. Usando esta información, el entrenamiento se realiza de manera iterativa, ajustando los parámetros de todas las neuronas en la RNA con el objetivo de aumentar su precisión al realizar predicciones sobre distintas entradas. Para esto, se busca reducir la diferencia entre las predicciones realizadas por la RNA y las etiquetas reales asociadas a cada entrada de ejemplo. El término iterativo se refiere a repetir el proceso de prueba y ajuste de parámetros hasta que la RNA sea capaz de mejorar sus predicciones.

6. Evaluar el modelo. Finalmente, es importante medir la capacidad de la RNA para identificar patrones de toxicidad en comentarios de texto que no se utilizaron como ejemplos durante su entrenamiento. La Figura 11 muestra este proceso. Dicha medición permite evaluar la calidad de las predicciones realizadas por la RNA. Con base en esa información, es posible determinar si la RNA es suficientemente robusta para utilizarse en un contexto real de aplicación, o si es necesario hacerle modificaciones para evitar comportamientos erráticos.

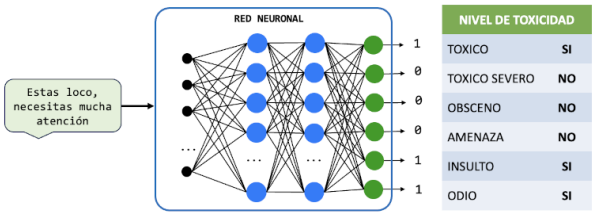


Figura 11. Evaluación de la RNA.

Conclusiones

Las plataformas digitales, y en especial las redes sociales, son espacios ampliamente utilizados por los usuarios de Internet para expresar sus opiniones y mantener comunicaciones interpersonales. Sin embargo, en estos espacios suelen difundirse comentarios tóxicos que causan daños a las personas y comunidades.

Realizar la detección de los niveles de toxicidad que contienen las expresiones que se dan en estas plataformas, es un desafío latente y complejo que se debe seguir abordando mediante el uso de las tecnologías más avanzadas, como los modelos de inteligencia artificial.

Las redes neuronales artificiales y sus diferentes arquitecturas son una opción prometedora que ha mostrado buenos resultados en la solución del problema de detección de textos tóxicos. Lo anterior se atribuye a su habilidad para identificar patrones complejos en textos y a su capacidad para adaptarse mediante el ajuste de sus parámetros. Por esto, se recomienda su uso en aplicaciones reales que permitan prevenir y manejar conductas dañinas en las plataformas digitales.✷

REFERENCIAS

1. Sheth, A., Shalin, V. L. y Kursuncu, U. (2022). Defining and detecting toxicity on social media: context and knowledge are key. *Neurocomputing*, 490, 312-318.

2. Madhyastha, P., Founta, A. y Specia, L. (2023). A study towards contextual understanding of toxicity in online conversations. *Natural Language Engineering*, 29(6), 1538-1560.

3. Hutto, C. y Gilber,t E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. En *Proceedings of the international AAAI conference on web and social media*, 8(1), 216-225.

4. Davidson, T., Warmsley, D., Macy, M. y Weber, I. (2017). Automated hate speech detection and the problem of offensive language. En *Proceedings of the international AAAI conference on web and social media*, 11(1), 512-515.

5. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., y Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2), 267-307.

6. Mehta, A. y Mehta, S. (2020). Detecting conversational toxicity using neural networks. *Social Science Research Network*.

7. Zhang, L., Wang, S. y Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1253.

8. Ahamed, K. I., y Akthar, S. (2016). A study on neural network architectures. *Computer Engineering and Intelligent Systems*, 7(9), 2222-1719.

9. Li, H., Mao, W. y Liu, H. (2019). Toxic comment detection and classification. En *CS299 Machine Learning*. Standford University.

10. Naidu, B. R., Tangudu, N., Sekhar, C., Kavitha, K., Ramana, B. V., Reddy, P., Sahukaru, J. y Lopinti, R. G. (2023). Toxic Comment Classification using Deep Learning. *International Journal On Recent And Innovation Trends In Computing And Communication*, 11(7), 93-104.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

SOBRE LOS AUTORES



Allan Balam Rueda Gutiérrez es Ingeniero en Sistemas Computacionales por la ESCOM-IPN, Maestro en Ciencias en Informática por la UPIICSA-IPN. Es docente en el Tecnológico Nacional de México, campus Tlalnepantla. Realiza investigación en el campo de Redes Adversarias Generativas.



Alejandro Rodríguez-Molina es Ingeniero en Sistemas Computacionales por la ESCOM-IPN, Maestro en Ciencias de la Computación por el CINVESTAV-IPN y Doctor en Ingeniería de Sistemas Robóticos y Mecatrónicos por el CIDETEC-IPN. Es profesor e investigador de la Universidad Autónoma de la Ciudad de México y realiza investigación en Inteligencia Artificial aplicada a la solución de problemas en ingeniería.



Oscar Armando Gómez Vargas es Ingeniero Físico por la UAM, Maestro en Ciencias con Esp. en Sistemas de Manufactura por el ITESM y Doctor en Ciencias en Ingeniería Mecánica por la ESIME-IPN. Es profesor investigador del Tecnológico Nacional de México, campus Tlalnepantla.



Eric Hernández Castillo es Ingeniero Electromecánico por el Instituto Tecnológico de Tlalnepantla, Maestro en Ciencias en Control Automático y Doctor en Ciencias en Control Automático, ambos por el CINVESTAV-IPN. Es subdirector académico del Tecnológico Nacional de México, campus Tlalnepantla.

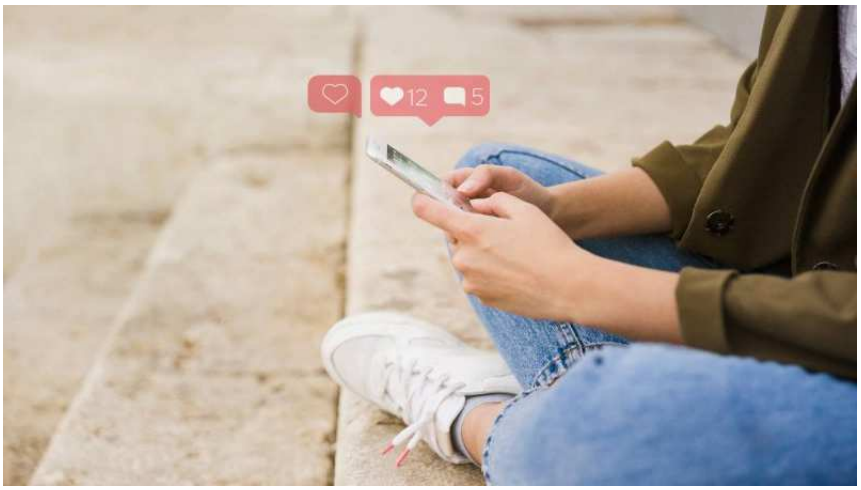


Imagen tomada de:
https://www.uautonoma.cl/noticias/toxicidad-digital-el-desafio-de-la-responsabilidad-en-redes-sociales/.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

ARTÍCULO ACEPTADO

Implementación práctica de Matlab® y visión estereoscópica para la medición de distancia de una persona

Rodrigo Hinojosa-Cruz, Juan Manuel Jacinto-Villegas y Otniel Portillo-Rodríguez

Introducción

En los seres humanos cada ojo proporciona una visión 2D de la perspectiva del mundo que nos rodea, es decir que, cada ojo percibe imágenes individuales [1]. Pero eso no es lo que nos hace diferentes a otros seres vivos, lo que verdaderamente nos permite obtener la percepción del mundo que tenemos es la combinación de estas dos perspectivas, en cuyo caso, el cerebro realiza una mezcla entre esas dos perspectivas para formar un par estereoscópico que va a contener información acerca de las tres dimensiones (percepción 2D y una profundidad Z).

Condiciones de la visión binocular y su campo visual

En los seres humanos, para obtener una visión binocular del mundo que nos rodea es necesario que se presenten las siguientes condiciones:

- Los campos visuales monoculares se superpongan en todas las direcciones de mirada.
- Los campos de fijación individuales se superpongan debido a los movimientos coordinados de los dos ojos.
- La transmisión neuronal desde los dos ojos alcance la misma área cerebral para que se produzca la fusión de las imágenes y se dé la coordinación de la percepción.

Para resumir lo anterior, en la Figura 1 se muestra el campo visual común de la visión binocular en los seres humanos; en ella, también se puede observar el campo visual en los ejes vertical y horizontal con los ángulos correspondientes para el campo de visión que se puede alcanzar en cada perspectiva.

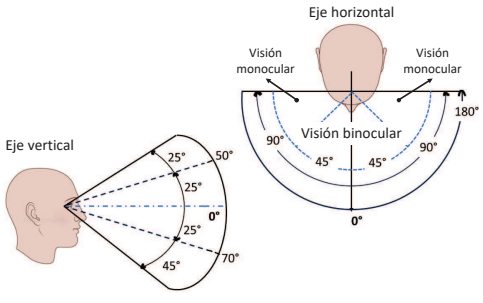


Figura 1. Esquema de campo visual común y la relación entre sus paredes y los ejes visuales.

Disparidad binocular

La percepción de la profundidad en Imágenes Generadas por Computadora (IGC) es considerablemente mejorada por la visualización estereoscópica. De esa forma, la proyección de IGC sobre la pantalla debe simular las características requeridas para implementar la visión estereoscópica[2].

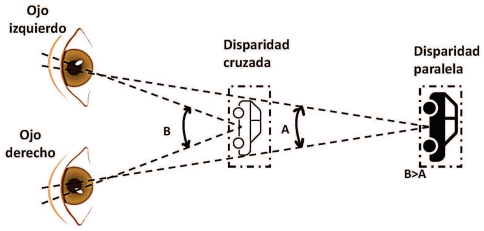


Figura 2. Disparidad binocular cruzada y paralela.

De lo anterior, la disparidad binocular se puede observar de dos formas: cruzada y paralela, y en cada una se va a percibir una distancia específica dependiendo el uso que se quiera dar (ver Figura 2).

En ese sentido, el uso de visión artificial (VA) permite replicar esta forma de percepción al llevar a cabo un procesamiento computacional de imágenes con el fin de detectar objetos, formas, personas y colores dentro de un espacio de trabajo, conocido también como campo de visión; el cual depende directamente de las propiedades de las cámaras que se usen así como las ubicaciones desde donde se realizan la captura de las imágenes.

Descripción de los parámetros intrínsecos y extrínsecos de una cámara

Para conocer las características del sensor de imagen de una cámara o videocámara, es necesario obtener los parámetros intrínsecos y extrínsecos.

- **Intrínsecos**.- estos están relacionados a las características internas de la cámara, como la distancia focal, el centro óptico (conocido como punto focal) y el coeficiente de inclinación, etc.
- **Extrínsecos**.- estos parámetros están relacionados con la localización (posición, orientación y traslación) de la cámara 2 con respecto a la cámara 1 en la escena 3D.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

De esta forma, para obtener un mejor procesamiento de video, hay tres factores que se deben tomar en cuenta:

- La resolución de la cámara tiene una alta influencia en la calidad de la grabación, así como también del procesamiento de visión por computadora.

  Resoluciones de 1280 px $*$ 720 px, 1920 px $*$ 1080 px, o superiores garantizarán una mejor adquisición.
- Los fotogramas por segundo (FPS) ideales para una buena visión deben ser superior a 24 FPS y de 60 FPS si se requieren escenas que implican objetos con movimientos rápidos.
- La velocidad de obturación afecta la claridad de los contornos del objeto en movimiento, algunas cámaras cambian esta velocidad dependiendo de la iluminación para mantener el mismo brillo.

En la actualidad, la implementación de VA nos permite incluso obtener la profundidad de diferentes objetivos, como por ejemplo: en [4] utilizan visión estereoscópica para la medición de la distancia de rostros en entornos industriales, de forma similar, en [5] presentan un enfoque de robótica colaborativa, en el que implementan este tipo de visión para lograr la medición de distancia de una persona mediante la detección del miembro superior y obtener la distancia a la que esta se encuentra de un robot, y en [6], aprovechan los datos de la cámara estéreo para la medición en tiempo real de obstáculos y/o personas para ser evitados por robots en entornos industriales.

En [7] desarrollaron un sistema de visión estereoscópica para el área de vialidad, este tiene el objetivo de medir la distancia de los peatones, y su principal contribución es el uso y optimización de filtros para eliminar objetos del entorno que no sean peatones. En [8], con el uso de este tipo de visión, plantean una nueva forma de abordar la medición y sensado en el área de salud, específicamente para personas invidentes, que en este caso se refiere a la medición de objetos para evitar colisiones. Por otro lado, relacionado con el área académica, en [9] hacen uso de la visión estereoscópica para detectar la temperatura de estudiantes y la distancia entre ellos para así predecir síntomas de enfermedades en las escuelas.

Los trabajos mencionados tienen en común la implementación de visión estereoscópica para la medición de distancia con respecto a diferentes objetivos. Sin embargo, no mencionan a detalle el procedimiento que utilizaron para adquirir los pares de imágenes y videos. En ese sentido, es importante resaltar que Matlab® es un software que dentro de sus herramientas de *visión por computadora y procesamiento de imágenes*, se encuentra una aplicación para la implementación de visión estéreo (stereo camera calibrator app). Sin embargo, no se brindan las herramientas suficientes para la captura de ambas imágenes (izquierda y derecha) y llevar a cabo su calibración, ni para la grabación de video contemplando dos perspectivas diferentes.

De acuerdo con lo anterior, la principal contribución de este artículo de divulgación se enfoca en brindar una guía y las herramientas necesarias para utilizar Matlab® de forma práctica y sencilla con visión estereoscópica, para realizar el procesamiento de videos y lograr la medición de la distancia a la que se encuentra una persona dentro de un espacio tridimensional.

Métodos y materiales

En esta sección se describe el software y hardware que fueron utilizados para el proyecto, así como también otros recursos.

Software

- Matlab® en su versión 2022a. Hasta el momento, versiones superiores no permiten la correcta implementación de visión estereoscópica.
- Herramienta de Matlab® para procesamiento de imagen y visión por computadora.
- Aplicación *stereo camera calibrator*.
- Scripts *snapShotStereo_Camera.m*, *video_test_winds_new.m* y *VideoStereo_Camera.m*.

Hardware

- Dos cámaras (web logitech c920).
- Tablero de calibración. Para este proyecto se diseño el tablero con dimensiones de 6x9 cuadros (simétrico en y pero asimétrico en x [10]) de 25 mm cada uno.
- PC con memoria RAM 16 Gb, procesador core i5.

Si se requiere más información y recursos del ejemplo práctico como: códigos de Matlab®, datos técnicos, entre otros, se recomienda consultar el siguiente repositorio de GitHub© clic aquí, que contiene dichos archivos y la descripción del procedimiento para la calibración e implementación de los scripts.

Metodología

Capturas de imágenes estéreo

El sistema estéreo de esta guía práctica consiste en dos cámaras con las mismas propiedades, las cuales están separadas a una distancia conocida (obtenida y almacenada en los parámetros extrínsecos) entre ellas, la relación entre sus ejes debe ser paralela o cruzada para tener un campo de visión binocular (ver Figura 3).

Debido a que se requiere eliminar imágenes con errores como distorsiones y desenfoque, en [3] se recomienda capturar un mínimo de 10 a 20 imágenes con cada una

de las cámaras (con perspectiva izquierda y derecha) para poder proceder a la calibración estéreo [10]. En este proyecto se generaron 30 pares de imágenes (ver Figura 4).



Figura 3. Posición frontal de las cámaras Logitech C920 HD PRO.

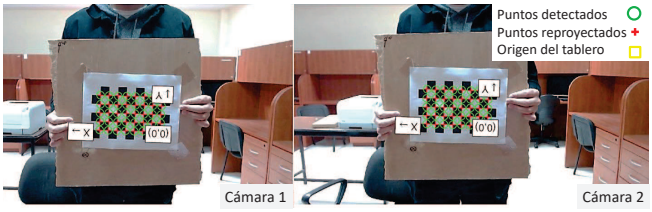


Figura 4. Captura de pares de imágenes izquierda y derecha con tablero de calibración.

Calibración estéreo de las cámaras

En esta sección se obtienen los parámetros intrínsecos y extrínsecos de las cámaras que se usan para realizar la captura de video (el objeto "stereoParams.m"). Posteriormente, en la Figura 5, se muestra el proceso que se sigue para ingresar y procesar las imágenes capturadas a la aplicación *Stereo Camera Calibrator App*.



Figura 5. Proceso para obtener los parámetros estéreo, implentando "Stereo Camera Calibrator"de Matlab.

Para este artículo se utilizaron 30 pares de imágenes, de las cuales 9 pares fueron descartados. Una vez cargadas las imágenes y tras eliminar las que representaban un error de proyección muy alto, se generaron los parámetros estéreo con un error absoluto de proyección medio de 0.21 px (ver Figura 6), siendo aceptable de acuerdo a la regla general en la que se contempla que los errores medios de reproyección de menos de un píxel son aceptables y también de acuerdo a la aplicación que se le da al sistema estereoscópico en este caso [3].

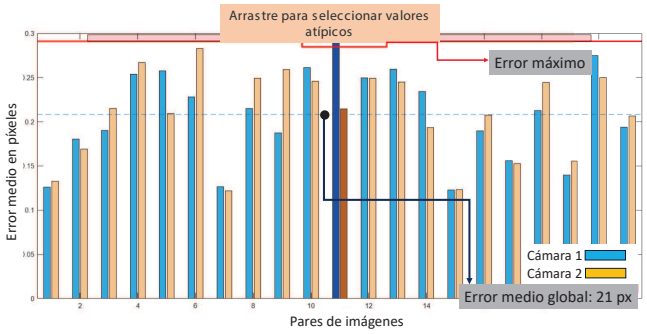


Figura 6. Errores de reproyección y error medio global.

En la Figura 7 se muestra una reconstrucción del campo de visión con planos tridimensionales donde se observa la posición/orientación de ambas cámaras, también se observan las posiciones/orientaciones del tablero de calibración que se usaron para la caracterización de las cámaras.

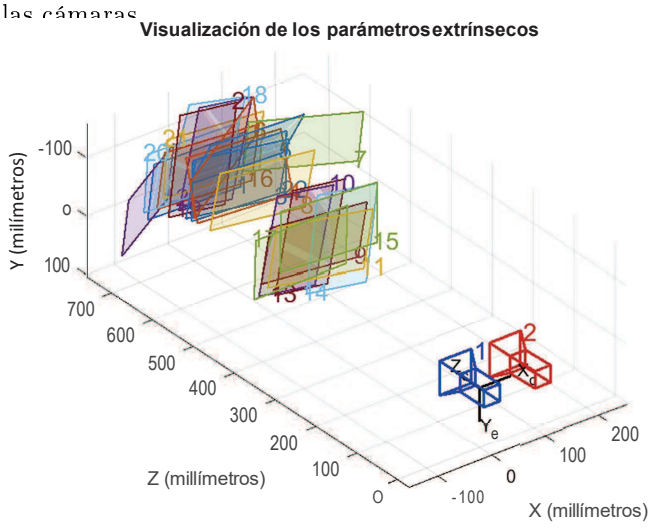


Figura 7. Visualización de los parámetros extrínsecos de las cámaras.

Grabación de video I&D (Izquierdo y Derecho)

Una vez obtenido los parámetros de visión estéreo, y sin cambiar de posición ambas cámaras, se procede a grabar los videos (de una persona) para cada perspectiva de forma simultánea. Cabe mencionar que Matlab no tiene una aplicación para captura de video en proyectos de visión estéreo (grabación de videos monoculares I&D simultáneamente) por lo que, para este proyecto se creó un script para grabar el video de cada perspectiva a una resolución de 1280 px $*$ 720 px y 25 FPS, siendo este un parámetro intrínseco correspondiente directamente al modelo de las cámaras usadas en este proyecto y a las características de procesamiento de la computadora donde se ejecutaron los scripts.

De esa forma, teniendo en cuenta las características intrínsecas de las cámaras, se debe configurar el

script para ajustar los valores y grabar los videos con perspectiva izquierda y derecha (I&D) con las mejores características de las cámaras. Cabe mencionar que el video se realizó inicialmente con el sujeto posicionado a una distancia de 400 cm de las cámaras, posteriormente 350 cm y 300 cm de distancia, que fueron medidos con un flexómetro.

Procesamiento de video para identificar distancia de personas

Una vez grabados los videos izquierdo y derecho, se debe proceder a incluir los archivos necesarios (StereoParams.m, Video_Left y Video_Right) al programa de captura de distancia *VideoStereo_Camera.m*. Así mismo, es importante mencionar que para evitar un error en la comparación de parámetros es necesario verificar que la resolución usada para la captura de imágenes izquierda y derecha en la calibración estéreo, debe ser la misma que se usa en la grabación de los videos.

Implementación de anaglifo, disparidad y escaneo 3d.

Después de ejecutar el script de Matlab® antes mencionado, y de acuerdo al proceso mostrado en la Figura 8, se abren ventanas emergentes mostrando procesos que se van aplicando a los videos para que posteriormente se realice la convergencia de las dos percepciones y poder calcular la distancia a la que se encuentra la persona en el video.



Figura 8. Bloques para el procesamiento de video.

El primer paso es la rectificación de los fotogramas de las cámaras izquierda y derecha que se realizan con el método de las líneas epipolares. La rectificación tiene como objetivo colocar ambas imágenes en el mismo plano, es decir, en su forma estándar de visión etéreo, posteriormente esta rectificación nos permitirá calcular la disparidad y reconstruir la escena 3D. Cabe mencionar que las imágenes rectificadas se pueden combinar en un anaglifo[1] (ver Figura 9), con el que se puede apreciar la percepción de profundidad utilizando lentes 3D rojo-azul (también llamados lentes anaglifos).

Posterior a la rectificación de cada uno de los fotogramas se debe realizar la computación de la disparidad. De forma general, la disparidad nos permite reconocer la distancia a la que se encuentran los objetos dentro de las escenas estudiadas.

[1] Los anaglifos se representan como una técnica de visualización de imágenes 3D en la que se muestran dos imágenes planas (estéreo par) superpuestas en una sola.



Figura 9. Imágenes rectificadas y utilizadas para representación 3D con anaglifo.

El procesamiento anterior se puede entender como el proceso en el que para cada píxel de la imagen izquierda se calcula la distancia al píxel correspondiente de la imagen de la derecha (cualquier par de puntos correspondientes se ubican en la misma fila de píxeles). En ese sentido, esa distancia denominada disparidad, es proporcional a la distancia del punto del mundo real desde la cámara (ver Figura 10).

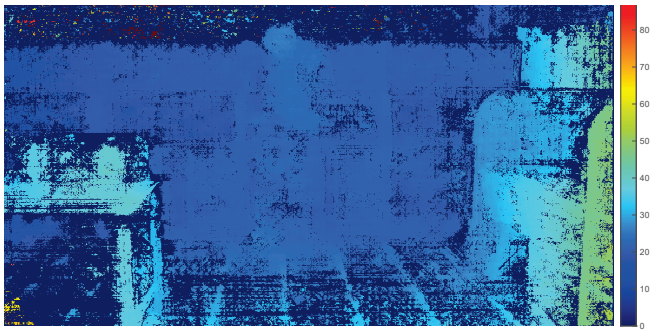


Figura 10. Obtención del mapa de disparidad con uso de colores para representar cercanía de los cuerpos.

En este punto, una vez que se obtuvo el mapa de disparidad, se puede generar una nube de puntos (ver Figura 11) que nos permite realizar una reconstrucción de la escena tridimensional de acuerdo a los puntos 3D correspondientes a cada píxel del mapa de disparidad.

Como última parte del procesamiento se debe reconocer y detectar a la persona en el video durante toda su reproducción con ayuda de la herramienta *vision.PeopleDetector* de Matlab®, con el objetivo de identificar a la persona que se encuentra dentro de la escena 3D de cada uno de los fotogramas estéreo del video, y además obtener el centroide (el punto de interés o coincidente) de la persona detectada.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

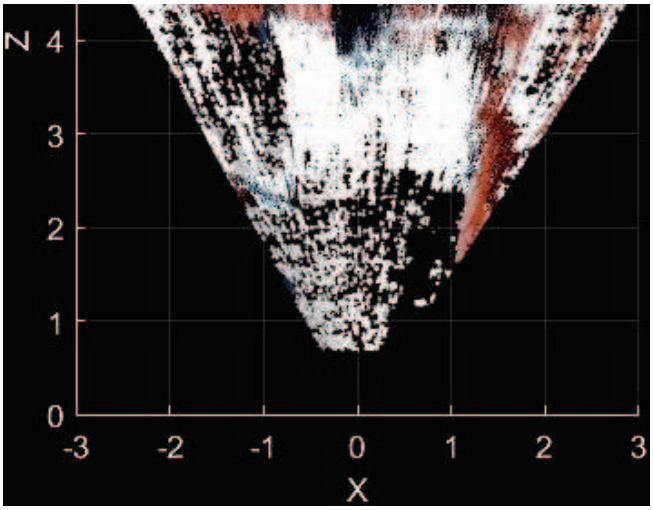


Figura 11. Nube de puntos para reconstrucción de escena 3D, vista desde los ejes X, Z.

Posteriormente, con el uso del mapa de disparidad se identifican las coordenadas tridimensionales del centroide de la persona detectada para calcular la distancia (proporcionada en una unidad de medida en metros) del sujeto desde la cámara.



Figura 12. Medición de profundidad de una persona.

Resultados y análisis

Como resultado de la implementación práctica de Matlab® y visión estereoscópica, en la reproducción del video, el sistema detectó al usuario y midió aproximadamente 402 centímetros de distancia en su primera posición (ver Figura 12), este proceso se replicó en cada uno de los fotogramas del video. De tal forma que, dentro de la etapa de experimentació, se presentaron diferentes mediciones con un error de variación menor a 20 cm con respecto a la distancia real a la que se encontraba la persona en las posiciones contempladas para las pruebas.

El error que se presenta en la medición se puede atribuir a factores como: la calibración de las cámaras, las dimensiones del tablero (una problemática durante el proyecto fue que el tamaño del tablero debe estar relacionado al tamaño del objetivo de estudio) y la cercanía del usuario a las cámaras, ya que ambas cámaras deben ver el cuerpo completo en todo momento para que el algoritmo (vision.PeopleDetector) reconozca correctamente el centroide de la persona detectada.

En la Tabla 1 se muestran tres comparaciones de los resultados de las distancias obtenidas con el sistema de visión estéreo y las reales.

Tabla 1. Comparaciones entre la posición real y la distancia obtenida por el sistema.

Posición real (cm)	Distancia obtenida (cm)
400	402
350	355
300	317

Conclusiones

En este artículo de divulgación se presentó una guía práctica, scripts y otros materiales para usar Matlab® y visión estereoscópica en un caso práctico para la detección y medición de distancia de una persona. Por lo tanto, el objetivo fue cumplido totalmente, logrando un procesamiento adecuado del video estéreo y obteniendo la medición de la distancia a la que se encuentra la persona en el espacio tridimensional reconstruido durante todo el video.

Como recurso extra, en este artículo se proporciona una forma práctica de visualizar los conceptos básicos de la implementación de visión estéreo, como pueden ser: una representación 3D de profundidad con un anaglifo, reconstruir el mapa de disparidad y una nube de puntos para la reconstrucción de escenas tridimensionales. Aunado a lo anterior, estos conceptos servirán como guía para diversos alumnos o como material para cursos de visión por computadora, así como de robótica en robots móviles y colaborativos.

Adicionalmente, para obtener un buen procesamiento de la imagen con visión estéreo se mencionan algunas sugerencias para configurar los scripts, como tener una alta resolución (1280∗720) px y FPS (superior a 24 FPS, o 60 FPS si se tienen escenas que implican objetos con movimientos rápidos) de ambas cámaras, realizar una buena calibración de los parámetros estéreo (el error de reproyección medio global debe ser menor a 1 px para ser aceptable) y, por último, que ambas cámaras observen el cuerpo completo del usuario en todo momento. ✷

REFERENCIAS

1. Álvaro, M., Pons-Moreno, F., Martinez-Verdú, R. 2004. Fundamentos de visión vinocular. Primera ed, Publicacions de la Universitat de València, ISBN 9788437093826, 408.

2. Poyade, M., y Lecuona, A. R. (2007). Análisis de la Disparidad Binocular en la Percepción de la Profundidad en Realidad Virtual. En Actas del VIII Congreso Internacional de Interacción Persona Ordenador (INTERACCIÓN 2007), 213-220.

3. MathWorks. (2015). Using the Stereo Camera Calibrator App. Recuperado el 10 de octubre de 2023 de `https://la.mathworks.com/help/vision/ug/using-the-stereo-camera-calibrator-app.html?searchHighlight=Stereo%20Calibration%20app%2C&s_tid=srchtitle_Stereo%20Calibration%20app%252C_2`.

4. Dandil, E., y ÇEVİK, K. K. 2019. Computer vision based distance measurement system using stereo camera view. En 2019 3rd *International Symposium on Multidisciplinary Studies and Innovative Technologies* (ISMSIT), 1-4.

5. Dewantara, B. S. B., Ardilla, F., y Thoriqy, A. A. 2019. Implementation of depth-HOG based human upper body detection on a mini PC using a low cost stereo camera. En *2019 International Conference of Artificial Intelligence and Information Technology* (ICAIIT), 458-463.

6. Eppenberger, T., Cesari, G., Dymczyk, M., Siegwart, R., y Dubé, R. 2020. Leveraging stereo-camera data for real-time dynamic obstacle detection and tracking. En *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS), 10528-10535.

7. Nguyen, U., Rottensteiner, F., y Heipke, C. (2019). Confidence-aware pedestrian tracking using a stereo camera. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, 53-60.

8. Bastomi, R., Ariatama, F. P., Putri, L. Y. A. T., Saputra, S. W., Maulana, M. R., Syai'in, M. y Zuliari, E. A. (2019). Object detection and distance estimation tool for blind people using convolutional methods with stereovision. En *2019 International Symposium on Electronics and Smart Devices* (ISESD), 1-5.

9. Soni, V. D., Soni, A. N., y Pothuganti, K. (2020). Student body temperature and physical distance management device in classroom using 3D stereoscopic distance measurement. *International Journal of Innovative Research in Science Engineering and Technology*, 9(9), 9294-9299.

10. MathWorks. (2015). Camera Calibration with Matlab. Recuperado el 10 de octubre de 2023 de `https://la.mathworks.com/help/vision/ug/using-the-stereo-camera-calibrator-app.html`.

SOBRE LOS AUTORES



Rodrigo Hinojosa Cruz obtuvo el grado de Ingeniero en Mecatrónica en 2021. Actualmente estudia la Maestría en Ciencias de la Ingeniería. Sus intereses científicos incluyen mecatrónica, robótica, visión por computadora, interacción humano-robot, automatización, interacción háptica.



Juan Manuel Jacinto Villegas obtuvo el grado de Ingeniero en Electrónica, y Maestro (con honores) en Ingeniería Electrónica, por el Instituto Tecnológico de Orizaba en 2010 y 2012, respectivamente. En 2017, obtuvo el grado de Doctor (con honores) en Tecnologías Digitales Emergentes en el área de investigación de Robótica Perceptual de la Scuola Superiore Sant'Anna, Pisa, Italia. Desde el 2018 es Profesor Investigador del programa "Investigadoras e Investigadores por México" del CONAHCYT, asignado a la Facultad de Ingeniería de la Universidad Autónoma del Estado de México (UAEM), Toluca, México. Sus áreas de intereses incluyen robótica, control háptico, interacción humano-robot, electrónica, mecatrónica y automatización.



Otniel Portillo Rodríguez obtuvo el grado en Ingeniería Electrónica en el Instituto Tecnológico de Toluca en el año 2000. En 2002, obtuvo el grado de Maestro en Sistemas Electrónicos por el Instituto Tecnológico de Estudios Superiores de Monterrey Campus Toluca. En 2008, le fue concedido el grado de Doctor en Robótica Perceptual por la Scuola Superiore Sant'Anna, Pisa, Italia. Actualmente, es profesor de tiempo completo asociado a las facultades de Ingeniería y Medicina de la Universidad Autónoma del Estado de México. Apoya a los programas de maestría y doctorado en ingeniería de dicha facultad (ambos pertenecientes al PNPC). En la facultad de medicina está asociado a la licenciatura en Bioingeniería Médica. Sus áreas de interés incluyen: machine learning, dispositivos mecatrónicos, procesamiento de señales e imágenes.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

ARTÍCULO ACEPTADO

Revelación de datos útiles para la industria turística de Guanajuato mediante Aprendizaje Profundo y Procesamiento del Lenguaje Natural

Angel Díaz Pacheco, Dora Luz Almanza Ojeda, Ansel Y. Rodríguez González, Miguel Á. Álvarez-Carmona, Ramón Aranda y Mario Alberto Ibarra Manzano

El turismo desempeña un papel crucial en la economía de México, y esto se manifiesta de manera significativa en el estado de Guanajuato donde se ha evidenciado un compromiso continuo mediante inversiones significativas en el desarrollo de este sector. En este contexto, se presenta una metodología integral para mejorar las técnicas de investigación en turismo, integrando herramientas de procesamiento del lenguaje natural y aprendizaje profundo. Al aprovechar las vastas cantidades de datos disponibles en internet, generados tanto por turistas como por entidades de gestión de destinos, este enfoque reduce la dependencia de encuestas costosas para la recopilación de información. Además, la implementación de métodos automatizados y la disponibilidad cotidiana de datos facilitan análisis periódicos, asegurando la identificación oportuna de tendencias y hallazgos relevantes.

Introducción

Según la Organización Mundial del Turismo, en 2021, la industria turística aportó aproximadamente 1.9 billones de dólares al producto interno bruto mundial [1]. La proliferación de las redes sociales ha transformado de manera sustancial esta próspera industria, brindando nuevas oportunidades para la recopilación de datos valiosos en el ámbito decisional. La investigación en este sector se presenta como un componente esencial para preservar la preferencia de los viajeros y elevar la calidad de los servicios. En México, las encuestas en el lugar de destino constituyen la columna vertebral de la investigación turística, por lo cual, este proceso se torna costoso en términos de tiempo y recursos financieros, limitando su frecuencia de realización.

Guanajuato destaca como uno de los principales destinos turísticos en México, atrayendo a una considerable cantidad de visitantes extranjeros, con alrededor de 319,394 registrados solo en el año 2021 [2]. Reconociendo su vasto potencial, la Secretaría de Turismo del estado ha desempeñado un papel activo en el fomento del desarrollo de este sector. En este contexto, la actualización de los métodos de investigación, tal como se propone en el presente documento, se perfila como una valiosa contribución para el perfeccionamiento continuo de la industria turística en la región.

Soluciones a problemas similares

En la última década, numerosos estudios en el ámbito turístico han recurrido a las técnicas de Procesamiento del Lenguaje Natural (PLN) con el propósito de recopilar información relevante y automatizar diversos procesos en la industria. Según el informe presentado por Álvarez-Carmona y su equipo en 2022, el PLN ha sido predominantemente utilizado en la construcción de sistemas de recomendación, análisis de sentimientos en reseñas y en la creación de chatbots [3].

Entre los estudios examinados, destaca el trabajo realizado por Guerrero-Rodríguez y colaboradores (2021), que se centró en el análisis de reseñas de establecimientos en Guanajuato publicadas en TripAdvisor®. Utilizando técnicas de Procesamiento del Lenguaje Natural (PLN), en particular una técnica llamada información mutua (MI), pudieron identificar las palabras más relevantes en estas reseñas con mayor poder discriminatorio para una tarea de clasificación. Los resultados obtenidos fueron significativos para identificar los aspectos más relevantes mencionados en las reseñas, especialmente en lo que respecta a las atracciones más destacadas de Guanajuato.

Destaca entre los trabajos de interés el realizado por Meira y su equipo en 2022, donde proponen una metodología innovadora para anticipar las calificaciones de atracciones turísticas a partir de las reseñas de los visitantes. Es crucial subrayar la presencia de un notable desbalance de clases en este tipo de problemáticas, motivo por el cual los investigadores adoptaron un enfoque basado en la detección de datos anómalos para potenciar la eficacia de su método.

Finalmente, Olmos-Martínez y colaboradores (2023) exploraron cómo se presenta Cancún a los turistas extranjeros a través de artículos de periódicos en línea. Para recopilar los datos, emplearon técnicas como el web scraping, y aplicaron métodos de Procesamiento del Lenguaje Natural (PLN) como la técnica de Latent Dirichlet Allocation (LDA) para identificar los temas discutidos más relevantes. LDA es una técnica estadística utilizada para el modelado de temas que permite detectar la presencia de tópicos en una colección de documentos. El estudio reveló que una proporción significativa de los artículos se centraba en las medidas sanitarias necesa-

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

rias para viajar durante la pandemia. Además, dada la creciente preocupación por la violencia y la amenaza del crimen organizado, muchos artículos advertían sobre los riesgos asociados con viajar a Cancún.

Enfoque propuesto

Para este estudio, se recopilaron datos de diversas fuentes con el fin de entender mejor a los actores clave en la generación de información turística. Por un lado, se analizaron datos de Organismos de Gestión de Destinos (OGD), como la Secretaría de Turismo y empresas hoteleras. Por otro lado, se examinó el Contenido Generado por el Usuario (CGU), es decir, la información compartida por los propios turistas en plataformas como Instagram y Twitter (recopilados entre enero y diciembre de 2023). En total, se revisaron más de 14,000 documentos relacionados con el CGU y más de 64,000 documentos de los OGDs.

Las información compartida por turistas es Contenido Generado por el Usuario (CGU). El contenido de los Organismos de Gestión de Destinos (OGD) promueve el destino para obtener ganancias.

Para analizar esta gran cantidad de datos, se utilizó una enfoque integral basado en métodos de aprendizaje profundo (método propuesto denominado MCAPT), que permite entender mejor la información recopilada. Para una mejor comprensión se presenta una descripción gráfica de la metodología propuesta en la Figura 1.

Se realizaron varios pasos para preparar los datos, como convertir todo el texto a minúsculas, eliminar palabras innecesarias y normalizar las palabras para simplificar el análisis. Luego, se utilizaron técnicas avanzadas de aprendizaje profundo para transformar el texto en representaciones vectoriales, que son como “huellas digitales"que capturan el significado de cada documento. Estas representaciones se usaron para agrupar los documentos en temas similares y realizar análisis de sentimientos para entender si los textos reflejaban actitudes positivas, negativas o neutras. Para aquellos lectores interesados, se recomienda explorar los trabajos de Floridi y Chiriatti (2020) y de Mikolov y colaboradores (2013).

Para identificar entidades específicas en el texto, como nombres de personas o lugares importantes, se utilizó una técnica llamada Reconocimiento de Entidades Nombradas (REN). Esta herramienta es como una especie de “lector"que busca y destaca automáticamente nombres importantes en un texto. Imagina que estás leyendo un artículo sobre turismo y de repente ves **“Cancún"** o **“Playa del Carmen"** resaltados en negrita; eso es gracias al REN. Esta técnica es muy útil en diferentes áreas de la inteligencia artificial, como en resúmenes de texto y en la traducción automática, ya que ayuda a identificar los nombres clave que necesitan ser preservados o traducidos con precisión.

En resumen, el estudio utilizó métodos avanzados de procesamiento de texto para analizar datos turísticos de diversas fuentes, lo que proporciona una comprensión más profunda de las percepciones y experiencias de los turistas.

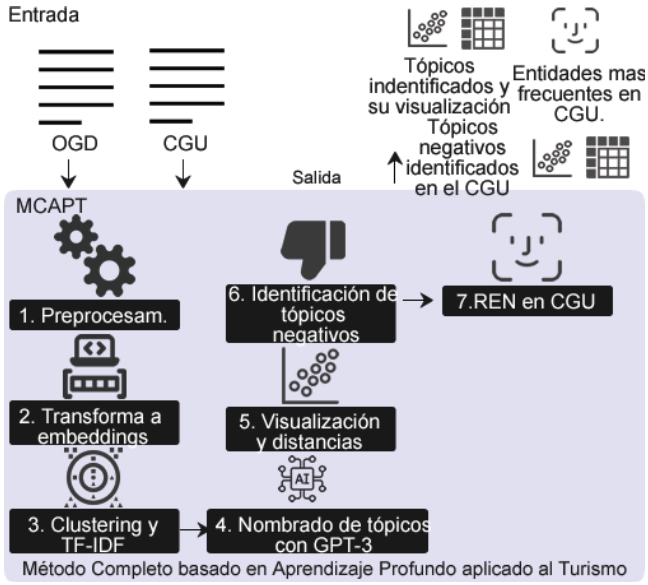


Figura 1. Descripción gráfica de la metodología integral propuesta.

Análisis de datos

Según el estudio de Gunn (2014), los turistas toman sus decisiones de viaje fundamentándose en la información que disponen sobre un destino específico. Por ende, es esencial realizar un análisis minucioso de la información, asegurando la clara distinción entre las fuentes de datos para mantener su integridad y confiabilidad.

Al examinar ambas fuentes de información (CGU y OGD), la metodología propuesta tiene la capacidad de proporcionar una visión integral que facilite una toma de decisiones informada. La Tabla 1 presenta una muestra de los tópicos identificados en ambas colecciones de datos para diversas plataformas y medios.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

En 2021 Guanajuato recibió 319,394 visitantes extranjeros.

Tabla 1. Palabras clave producto del análisis de tópicos y frases generadas a partir de las mismas con ChatGPT. Las palabras clave van sin tildes ni signos de puntuación debido al preprocesamiento.

Palabras clave	Frase
Contenido de las OGD	
Guanajuato, guero, alvarez, leon, dia, grandes, historias	Vida y legado del Guero Alvarez en Guanajuato
turismo, turistico, viajero, taller, desarrollo, sector, guanajuato	Desarollo turístico y talleres en Guanajuato
cultural, evento, noviembre, miguel, allende, festival, tradiciones	Celebrando las tradiciones y la cultura en San Miguel de Allende
programa, grande festival, cervantino, teatro	Gran programa del festival Cervantino en los teatros de Guanajuato
Contenido Generado por el Usuario	
alameda, buscamos, salud, ciclista	Promoción de la salud y el ciclismo en la alameda
obras, vivienda, proyecto, noticias, bajio, corrupción, gobierno	Planes de gobierno, desarrollo y corrupción en Guanajuato
ciudad, feria, leon, huauchinango, hidalgo, aniversario, celebración	Celebrando el aniversario de Leon Guanajuato, explorando la feria y festividades de la ciudad.
visita, si, museo, momias, guanajuato, mexico	Visita el museo de las momias en Guanajuato.

Dado que las representaciones generadas por el aprendizaje profundo actúan como huellas digitales de los documentos, podemos aprovechar todas las herramientas matemáticas disponibles para realizar análisis diversos, como calcular la similitud entre los discursos predominantes en el Contenido Generado por el Usuario y el contenido de los Organismos de Gestión de Destinos. Esta información es crucial, ya que nos proporciona pistas sobre si las estrategias de promoción del destino están logrando captar el interés de los turistas en las actividades propuestas. Aunque estos datos pueden expresarse como números, presentar la similitud visualmente, mediante puntos en un espacio tridimensional, facilita su comprensión. En la Figura 2, se muestran algunos de los temas mencionados en la tabla anterior. Los puntos azules corresponden a los temas identificados en los documentos de los OGD, mientras que los puntos rojos representan el Contenido Generado por el Usuario.

La Figura 2 muestra que los temas surgidos tanto del contenido de los Organismos de Gestión de Destinos como del material compartido por los turistas exhiben una notable cercanía en el espacio, lo que sugiere que las percepciones de los viajeros y las organizaciones no difieren significativamente. Es decir, los turistas muestran interés en lo que las OGDs están promocionando. Sin embargo, al identificar discrepancias entre los temas abordados por cada actor, se resalta la importancia de mejorar las estrategias de promoción de las OGDs. Este ajuste busca alinear el contenido compartido por los turistas con los objetivos promocionales establecidos por las organizaciones de gestión de destinos. En particular, se observa una divergencia destacada en la promoción de un festival cultural en San Miguel de Allende, representado como el punto más alejado del centro del diagrama, en contraste con la falta de menciones de dicho tema en los comentarios de los turistas.

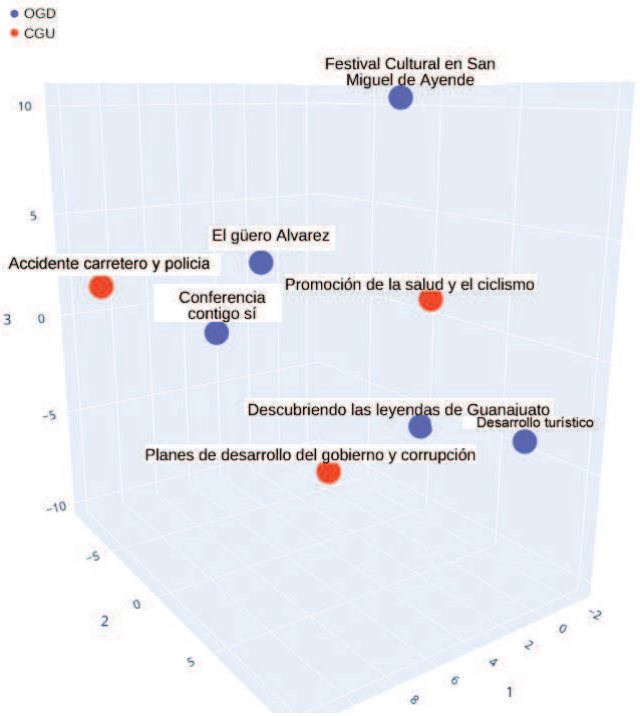


Figura 2. Representación espacial de los tópicos subyacentes en las colecciones de datos.

A continuación, se realizó un análisis de sentimientos en la sección correspondiente al Contenido Generado por el Usuario para identificar las principales quejas o deficiencias en la gestión del destino turístico (ver Tabla 2). Se identificaron claramente seis temas: inseguridad, violencia de género, higiene, el museo de las momias, la falta de guías capacitados y la gestión deficiente de atracciones turísticas debido a las largas filas. La problemática de seguridad en el estado de Guanajuato es ampliamente reconocida, y nuestros hallazgos simplemente respaldan la percepción general entre los ciudadanos.

Por otro lado, se observa una clara tendencia relacionada con los problemas en la gestión de servicios y atracciones, como lo demuestran las quejas sobre malos olores y basura en las calles, así como las largas filas para acceder a atracciones como museos y recorridos turísticos. También se resalta la importancia para los turistas de mejorar la calidad de la capacitación de los guías turísticos para optimizar su experiencia.

Finalmente, el museo de las momias genera controversia entre una parte significativa de los viajeros, ya que

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

algunos lo consideran un espectáculo morboso al exhibir cadáveres sin mostrar respeto. Sin embargo, este museo tiene un valor histórico considerable para la ciudad. Por lo tanto, es necesario abordar esta controversia de manera cuidadosa y exhaustiva para mantener la preferencia de los turistas y preservar la identidad histórica del lugar.

Tabla 2. Principales tópicos clasificados como negativos en el CGU y frases generadas en base a las palabras clave con ChatGPT. Las palabras clave van sin tildes ni signos de puntuación debido al preprocesamiento.

Palabras clave	Frase
guanajuato, palma, viajero, poder, hoy, utilizar, vamos, sierras, peor, robo	La prevención del crimen y robos en Guanajuato y otras áreas montañosas y la importancia de tomar medidas de seguridad adecuadas para garantizar la seguridad y protección de los viajeros en dichas áreas.
violencia, igualdad, queremos, mujeres, impulsamos	La guerra contra la violencia de genero y la promoción de la igualdad, especialmente en lo concerniente a la defensa de los derechos de las mujeres y como un cambio positivo en la sociedad puede empoderar y promover a través de diferentes promociones y estrategias.
huele, turista, basura, mantenimiento, vista, turístico, limpieza, sucio	La importancia del mantenimiento y la limpieza en los lugares turísticos para mejorar la vistas y prevenir a los turistas de percibir los malos olores a causa de la basura.
cadaveres, museo, horrible, momias, momia, peor, terrible, mala, muertas, mal	El controversial museo de las momias de Guanajuato, donde las cadáveres momificados y las personas que murieron hace muchos años, ha atraído importantes críticas como una práctica cuestionable.
decepcion, espera, mediocre, larga, fila, decepcionante, esperaba, hora, esperar, intentamos	La decepción y la frustración de esperar largas horas en filas, solo para tener experiencias mediocres
explicaciones, narradores, interes, historico, respeto, conocerlo, entender	La importancia de tener guías turísticos bien capacitados, para poder entender y aprecia mejor la historia y la cultura de un lugar puede incrementar el interés y la satisfacción de los turistas.



Figura 3. Nube de palabras con las atracciones más populares en el CGU. El tamaño de la palabra esta relacionada a su frecuencia en el contenido.

En cuanto al análisis de entidades nombradas, se aplicó para identificar las atracciones más populares entre los viajeros. Con el fin de hacer más accesible la interpretación de los resultados, se generó una nube de palabras (ver Figura 3), donde el tamaño de cada palabra refleja la frecuencia de menciones en el corpus. Es destacable que el Teatro Juárez figura como la atracción más mencionada por los visitantes, y es notable que todas las opiniones analizadas fueron positivas. Además, tanto la Alhóndiga de Granaditas como el Jardín de la Unión emergen como puntos destacados entre los encantos de Guanajuato.

Por otro lado, el museo de las momias y el callejón del beso son temas recurrentes en las conversaciones, principalmente debido a las deficiencias mencionadas anteriormente para el museo y a las multitudes y largas filas que enfrentan los visitantes para acceder al callejón.

Conclusiones

Este estudio presenta una metodología integral destinada al análisis exhaustivo de extensas colecciones de contenido textual, con el objetivo de extraer información valiosa para la toma de decisiones, considerando dos perspectivas fundamentales: la de los turistas (CGU) y la narrativa proporcionada por los organismos de gestión de destino (OGD). La aplicación de técnicas avanzadas de procesamiento del lenguaje natural permite el análisis de datos provenientes de diversas fuentes, abordando su naturaleza ruidosa y desordenada.

La metodología presentada tiene la capacidad de identificar los principales tópicos discutidos, los problemas más relevantes y las atracciones destacadas, así como de discernir las diferencias entre el discurso del contenido compartido por los viajeros y el promovido por las organizaciones.

Aunque el método presenta ventajas significativas, se reconoce como un trabajo en progreso que continuará su proceso de mejora con el fin de aumentar su utilidad y nivel de automatización. En el contexto de futuras mejoras, se contempla el desarrollo de una interfaz gráfica para facilitar su utilización por parte de usuarios no familiarizados con lenguajes de programación. Además, para abordar el análisis temporal de este fenómeno, será necesario extender este enfoque y emplear otros métodos para la extracción de información pertinente.✷

REFERENCIAS

1. World Tourism Organization (UNWTO). (2021). The Economic Contribution of Tourism and the Impact of COVID-19.

2. SECTUR-GTO. Observatorio Turistico del Estado de Guanajuato, March 2023.

3. Álvarez-Carmona, M.A., Aranda, R., Rodríguez-Gonzalez, A. Y., Fajardo-Delgado, D., Sánchez, M. G., Pérez-Espinosa, H., Martínez-Miranda, J., Guerrero-Rodríguez, R., Bustio-Martínez, L., & Díaz-Pacheco, Á. (2022). Natural language processing applied to tourism research: A systematic review and future research directions. En *Journal of King Saud University - Computer and Information Sciences*, 34(10), 10125–10144.

4. Guerrero-Rodriguez, R., Álvarez-Carmona, M. Á., Aranda, R., & López-Monroy, A. P. (2021). Studying Online Travel Reviews related to tourist attractions using NLP methods: the case of Guanajuato, Mexico. En *Current Issues in Tourism*, 26(2), 289–304).

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

5. Meira, J., Carneiro, J., Bol on-Canedo, V., Alonso-Betanzos, A., Novais, P., and Marreiros, G. (2022). Anomaly detection on natural language processing to improve predictions on tourist preferences. *Electronics*, 11(5):779.

6. Olmos-Martínez, E. , Álvarez-Carmona, M.A., Aranda, R., and Díaz-Pacheco, (2023). A. What does the media tell us about a destination? the Cancún case, seen from the USA, canada, and mexico. *International Journal of Tourism Cities*.

7. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. y Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.

8. Floridi, L. y Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30, 681-694.

9. Sharnagat, R. (2014). Named entity recognition: A literature survey. *Center For Indian Language Technology*, 1-27.

10. Gunn, C. A. (2014). Vacationscape: Developing tourist areas. *Routledge*.

SOBRE LOS AUTORES



Angel Díaz Pacheco. SNI Nivel I. Recibió en 2010 el grado de Ingeniero en Sistemas Computacionales, en 2013 el grado de Maestro en Sistemas Computacionales y en 2019 el grado de Doctor en Ciencias en el área de Ciencias de la Computación por el INAOE. Realizó una estancia posdoctoral en la Unidad de Trasferencia Tecnológica del CICESE. Actualmente es Profesor de tiempo completo en la Facultad de Ingenierías de la Universidad de Guanajuato. Sus intereses de investigación incluyen el aprendizaje automático, *big data*, el computo suave y la aplicación de las anteriores en la investigación turística.



Dora Luz Almanza Ojeda Es Ingeniera en Comunicaciones y Electrónica y Maestra en Ingeniería Eléctrica, por la Universidad de Guanajuato, Doctora en Sistemas Informáticos, por la Universidad de Toulouse III: Paul Sabatier, en Toulouse, Francia. Sus áreas de interés son la visión por computadora y el desarrollo de sistemas autónomos. Ha publicado 26 artículos en revistas de alto impacto, ha dirigido 27 trabajos de tesis entre Licenciatura, Maestría y 3 tesis de Doctorado. Es miembro del Sistema Nacional de Investigadores Nivel 1. Actualmente, es profesor-investigador en la Universidad de Guanajuato Sede Salamanca.



Ansel Y. Rodríguez González. SNI Nivel I. Recibió los grados de Licenciado en Ciencias de la Computación en 2004 y Maestro en Ciencias Matemáticas en 2007 en la Universidad de la Habana, Cuba, y el grado de Doctor en Ciencias Computacionales en 2011 en el INAOE. Desde 2018 es Investigador por México en CICESE-UT3. Sus intereses de investigación incluyen minería de datos, reconocimiento de patrones, aprendizaje por refuerzo y algoritmos evolutivos.



Ramón Aranda SNI Nivel I. Recibió en 2008 el grado de Licenciado en Ciencias de la Computación por parte de la Facultad de Matemáticas de la UADY, en 2010 recibió el grado de Maestro en Ciencias de la Computación con especialidad en Matemáticas Industriales y en 2016 el grado de Doctor en Ciencias con Orientación en Ciencias de la Computación, ambos posgrados por el CIMAT. De 2019 a la fecha, es Investigador por México (IxM - CONACyT), actualmente asignado al CIMAT, Unidad Mérida.



Miguel Á. Álvarez-Carmona. SNI I. Recibió en 2012 el grado de Ingeniero en Ciencias Computacionales por por la BUAP, en 2014 y 2019 el grado de Maestro y Doctor en Ciencias Computacionales por el INAOE donde se especializó en inteligencia artificial y procesamiento de lenguaje natural. Ha sido autor de diversos artículos en revistas y congresos nacionales e internacionales. De 2019 a la fecha, es investigador por México de CONAHCyT en CIMAT mty. Pertenece a la Asociación Mexicana de Procesamiento de Lenguaje Natural y a la Sociedad Mexicana de Inteligencia Artificial.



Mario Alberto Ibarra Manzano Es profesor titular en el Departamento de Ingeniería Electrónica de la Universidad de Guanajuato, con un doctorado en Microelectrónica y Microsistemas de la INSA, Toulouse, Francia. Investigador y académico destacado, ha publicado extensamente, incluyendo 44 artículos indexados en JCR, 5 capítulos de libros y más de 50 ponencias en conferencias, y tiene una patente nacional. Ha liderado múltiples proyectos de investigación y ha sido mentor de más de 70 estudiantes de tesis. Sus contribuciones a la telemática y la ingeniería electrónica son reconocidas por su inclusión en el Sistema Nacional de Investigadores (SNI-I) desde 2015.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

ARTÍCULO ACEPTADO

Modelado dinámico/identificación de robots usando redes neuronales artificiales: un enfoque para principiantes

Jesús A. Medrano-Hermosillo, Abraham E. Rodríguez-Mata, Larbi Djilali, Victor Alejandro Gonzalez Huitron y Oscar J. Suarez Sierra

Introducción

Desde la Revolución Industrial, se han producido innumerables modificaciones en diversos sectores laborales a nivel mundial. Con el transcurso del tiempo las automatizaciones han adquirido una relevancia cada vez mayor. En el ámbito específico de los sistemas robóticos, no sólo han contribuido al incremento de la producción, sino que también su integración en la industria ha experimentado un constante crecimiento a lo largo de los años. Sus aplicaciones van desde soldado, pintura, corte, recolección de objetos, ensamble, etcétera. Sin embargo, cada vez son más utilizados no sólo en el aspecto industrial, sino en otras áreas como en medicina, deportes, educación, entre otras.

Para ilustrar el crecimiento de los robots, la Figura 1 muestra el aumento significativo de los sistemas robóticos en la industria desde el año 2000 hasta 2016. Es relevante señalar que se observa una disminución en 2008, atribuida a la recesión económica de ese período. Sin embargo, según indican los autores en [1], se proyecta que el número de robots a nivel mundial se incremente en los próximos 20 años, alcanzando los 20 millones para el año 2030, de los cuales se estima que 14 millones estarán operando únicamente en China. Esto demuestra la importancia de estos sistemas en la industria del futuro.

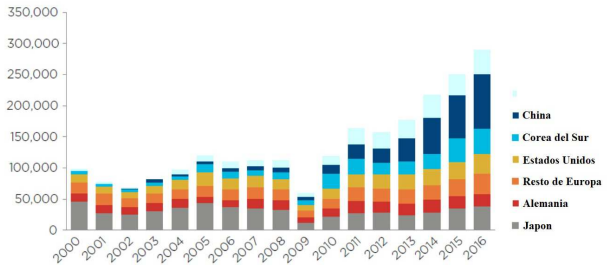


Figura 1. Instalación de sistemas robóticos entre los años 2000 al 2016 [1].

Si se habla del diseño específico de un sistema robótico, es necesario mencionar diferentes técnicas para su elaboración las cuales, de manera general, se describen a continuación [2,3]:

- Análisis cinemático. Este estudio se centra en el análisis del movimiento del robot sin tener en cuenta las fuerzas que lo influyen.
- Análisis dinámico. Este estudio se centra en analizar el sistema robótico desde la perspectiva de las fuerzas que actúan sobre él, es decir, representarlo mediante ecuaciones que reflejen su comportamiento en el mundo real.
- Leyes de control. Partiendo del análisis dinámico, el propósito es asegurar que el robot pueda seguir con precisión las instrucciones del usuario. Para alcanzar este objetivo, es necesario implementar una ley de control para seguir con precisión las trayectorias deseadas. Un ejemplo común de controlador es el PID (Proporcional, Integral, Derivativo).
- Planeación de trayectorias. Este análisis genera los valores angulares necesarios de un robot para seguir una ruta específica. Existen numerosas técnicas para lograr este objetivo, entre las cuales se incluyen SLAM (Localización y Modelado Simultáneos), algoritmos geométricos, campos potenciales artificiales, métodos optimizados, entre otros.

Con base en lo anterior, se observa una amplia variedad de herramientas para desarrollar un robot, siendo la elección más adecuada aquella determinada por la aplicación específica y el entorno de implementación. Sin embargo, un aspecto crucial es el uso del modelo dinámico del sistema para comprender su comportamiento en el mundo real y aplicar técnicas de control que aseguren su correcto funcionamiento. Por lo tanto, es indispensable calcular correctamente el modelo dinámico para que el control sea efectivo. De lo contrario, podrían surgir acciones no deseadas durante la aplicación en tiempo real, así como errores en las trayectorias planificadas. En resumen, un cálculo incorrecto del modelo dinámico puede afectar significativamente el funcionamiento adecuado de la aplicación.

Existen varias opciones para calcular el modelo dinámico de un robot, desde enfoques convencionales hasta técnicas avanzadas. Entre las técnicas tradicionales, destacan el modelado mediante Euler-Lagrange y el de Newton-Euler. Euler-Lagrange es eficaz para resolver

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

problemas de optimización y es fundamental en aplicaciones matemáticas y físicas, donde en sistemas electromecánicos describe su comportamiento dinámico. Por otro lado, Newton-Euler es un algoritmo iterativo que se basa en las leyes de Newton. Es importante destacar que ambas técnicas convergen en el mismo resultado. También existen métodos avanzados, como los geométricos. Sin embargo, al utilizar estas técnicas es crucial comprender todos los detalles del robot, ya que perturbaciones o parámetros inciertos llevan a un modelo erróneo. Además, el deterioro o las tolerancias afectan al modelo, haciendo necesario el uso de metodologías de identificación adaptativas, como las redes neuronales artificiales, que garantizan un modelo dinámico preciso y actualizado. Esto contrasta con los métodos tradicionales, que pierden precisión con el tiempo debido a cambios en las condiciones del robot [2,4].

Por lo tanto, se plantea la pregunta, ¿Qué es la identificación de un sistema dinámico? Es un algoritmo que aprende el comportamiento de un sistema dinámico, como un robot, a partir de ejemplos. La identificación de un robot implica encontrar los parámetros de su ecuación diferencial, generalmente linealmente independientes, que se pueden estimar con una red neuronal. El objetivo de este trabajo es identificar el modelo dinámico de un robot usando redes neuronales, las cuales proporcionan un modelo más preciso frente a perturbaciones y parámetros inciertos, en comparación con las técnicas tradicionales [5].

Redes Neuronales Artificiales

La red neuronal artificial (RNA) es un algoritmo inspirado en el cerebro humano, diseñado para emular el aprendizaje neuronal. La Figura 2 compara una neurona cerebral y una artificial con sus componentes principales, descritos a continuación [6]:

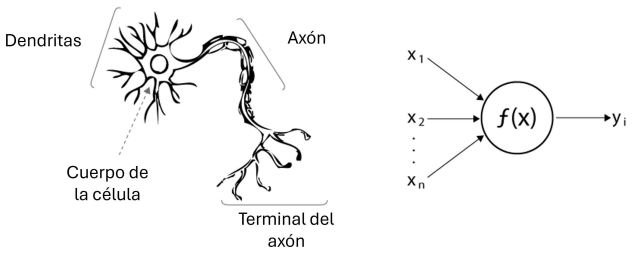


Figura 2. Relación entre una neurona cerebral y una artífical.

- Dendritas. Las estructuras ramificadas que se extienden desde el cuerpo celular se conocen como dendritas. Se utilizan para recibir señales de otras neuronas en forma de impulsos eléctricos, las cuales se transmiten al cuerpo celular para su procesamiento.
- Cuerpo de la célula. Es la parte que contiene el núcleo y la mayor parte de las estructuras celulares, como el citoplasma y los orgánulos. En las neuronas, el cuerpo celular integra las señales recibidas de otras neuronas a través de las dendritas y, en función de esta información, genera señales a través del axón para comunicarse con otras células.
- Axón. Son largas proyecciones que transmiten impulsos eléctricos lejos del cuerpo celular. Son esenciales en el sistema nervioso para enviar señales a otras neuronas, músculos o glándulas. En sus extremos, se encuentran las terminaciones axónicas, que se conectan con otras células para transmitir señales a través de sinapsis.
- Terminal del axón. Es la estructura final de una neurona, donde es el punto de contacto funcional entre una neurona y una célula diana, que puede ser otra neurona, una célula muscular o una célula glandular. Cuando una señal eléctrica viaja a lo largo del axón hasta llegar al terminal axónico, las vesículas sinápticas liberan neurotransmisores en el espacio sináptico, permitiendo así la comunicación entre neuronas.

Por su parte, la RNA incorpora elementos similares a los mencionados. Las entradas de la red neuronal artificial, representadas por x, reflejan las señales que reciben las dendritas. El equivalente al cuerpo de la célula en la RNA sería la agregación de todas las entradas de las señales de entrada que a su vez se multiplican por unas ganancias llamadas pesos w (las cuales las describiremos más adelante), es decir, esta etapa recopila toda la información proveniente de las señales de las dendrítas. Sin embargo, en una red neuronal artificial convencional, a menudo se aplica una función a esta suma, con el propósito de limitar la amplitud de la señal y evitar que se vuelva demasiado grande, es decir se expresaría de la siguiente manera $f(w_1x_1 + w_3x_2 + w_3x_3...)$. Aunque en ocasiones se emplea simplemente la suma directa de las señales de entrada $(w_1x_1 + w_3x_2 + w_3x_3...)$.

Finalmente, la salida de la red neuronal artificial es lo obtenido por el cuerpo de la célula, es decir, se representaría como $y = f(w_1x_1 + w_3x_2 + w_3x_3...)$, análogo a las señales de salida del axón y su terminal en una neurona cerebral.

Por otro lado, en el cerebro humano existe una red de millones de neuronas entrelazadas para transmitir y procesar información, desempeñando una amplia gama de tareas. Esta compleja interconexión inspira el diseño de Redes Neuronales Artificiales (RNAs), que simulan el proceso de aprendizaje cerebral al interconectar neuronas artificiales. La Figura 3 ilustra cómo la sinapsis entre las neuronas del cerebro se refleja en las redes neuronales artificiales. En esencia, al unir múltiples neuronas

en configuraciones seriales y/o paralelas, creamos redes neuronales artificiales que emulan el sistema cerebral. La organización de estas redes se denomina arquitectura, que puede variar ampliamente según la aplicación. Esta arquitectura suele dividirse en número de neuronas artificiales y capas, donde las capas son determinadas por su posición: la capa de entrada recibe los valores iniciales, la capa de salida produce los resultados finales, y las capas intermedias entre ellas se conocen como capas ocultas. En las capas ocultas, las entradas de la misma estarán conformadas por salidas de neuronas artificiales anteriores con sus respectivos pesos. En particular, cuando las capas ocultas son un número considerable, al algoritmo completo se le suele llamar aprendizaje profundo (*deep learning* por su traducción al inglés) [5,6].

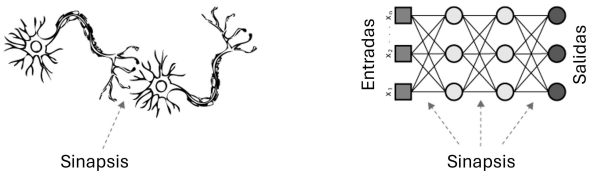


Figura 3. Analogías entre las interconexiones de redes cerebrales y artificiales.

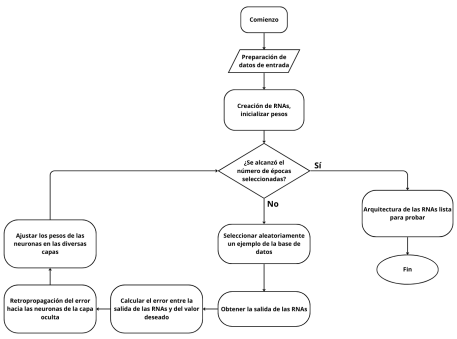


Figura 4. Algoritmo para el aprendizaje de RNAs fuera de línea.

Ahora que comprendemos la estructura de nuestras RNAs, surge la pregunta crucial: ¿cómo aprenden? En esencia, como nosotros lo hacemos. Necesitan ejemplos para ajustar su salida y minimizar la discrepancia entre lo esperado y lo obtenido. Esto implica proporcionar una amplia gama de ejemplos y repetir el proceso varias veces, al proceso de repetir continuamente los ejemplos se le denomina época. En resumen, iniciamos el algoritmo de aprendizaje proporcionando todos los ejemplos, y cada vez que volvemos a presentar el conjunto inicial de ejemplos, consideramos que ha pasado una época. El número de épocas necesario para que las RNAs aprendan de manera efectiva varía según la aplicación y la arquitectura del modelo.

Uno de los algoritmos que se usa para minimizar el error en una RNA se conoce como gradiente descendente. Esta técnica actualiza las ganancias, las cuales se multiplican por las entradas, llamadas pesos, para reducir el error entre la salida de la RNA y la salida esperada. Los pesos se ajustan de manera iterativa durante el entrenamiento, estabilizándose cuando el error se minimiza en la medida de lo posible para la aplicación específica. Cuando se trata de una sola neurona, es fácil determinar su error y aplicar el algoritmo de aprendizaje. Sin embargo, en redes neuronales con múltiples capas surge un problema, el cual es que las neuronas en las capas ocultas no tienen una señal deseada, lo que hace imposible calcular el error entre ellas. Por lo tanto, el algoritmo de gradiente descendente no es aplicable en estas circunstancias. Para resolver esta limitación, se recurre a otro algoritmo conocido como retropropagación del error (*backpropagation* por su traducción al inglés), que propaga el error desde la salida hacia las capas ocultas, permitiendo así ajustar los pesos y minimizar el error en la salida del sistema. El algoritmo mencionado se puede ver en la Figura 4, el cual se ilustra mediante un diagrama a flujo [6].

Identificación del modelo dinámico utilizando RNAs

Para realizar la identificación, es fundamental que la salida de las RNAs se asemeje estrechamente a la salida del robot [7]. Una vez que hemos reducido el error al mínimo posible, nuestras RNAs reflejarán el comportamiento del robot, y así habremos resuelto la función de identificación. El estrechamiento mencionado se alcanza al minimizar el error entre la respuesta del robot a un estímulo específico y la respuesta obtenida al aplicar ese mismo estímulo a las RNAs. Este proceso de minimización del error es dirigido por el algoritmo de entrenamiento, el cual ajusta continuamente los pesos de las RNAs en función de las salidas obtenidas, buscando reducir el error al máximo posible. Esta adaptación constante del sistema neuronal artificial permite hacer frente a cambios dinámicos futuros, como perturbaciones o tolerancias.

En la parte del aprendizaje, una práctica común para las RNAs es elegir épocas preestablecidas para un entrenamiento fuera de línea, donde la red se entrena por separado y luego se utilizan los pesos resultantes (al terminar las épocas), para aplicarlos en una tarea específica. Sin embargo, para identificar sistemas dinámicos en tiempo real, es fundamental emplear el entrenamiento en línea. Esto implica que el sistema permanece en constante actualización, siempre aprendiendo y adaptándose a medida que se reciben nuevos datos. En otras palabras, el algoritmo de aprendizaje está constantemente activo en estas aplicaciones. En la Figura 5 se observa un diagra-

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

ma de cómo se representa el algoritmo de identificación para un sistema robótico [8].

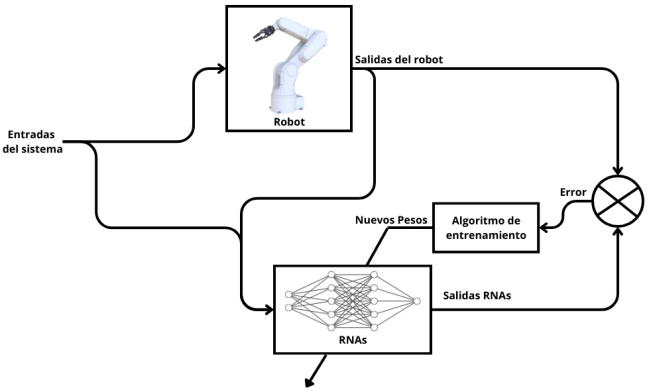


Figura 5. Algoritmo de identificación para un brazo robótico.

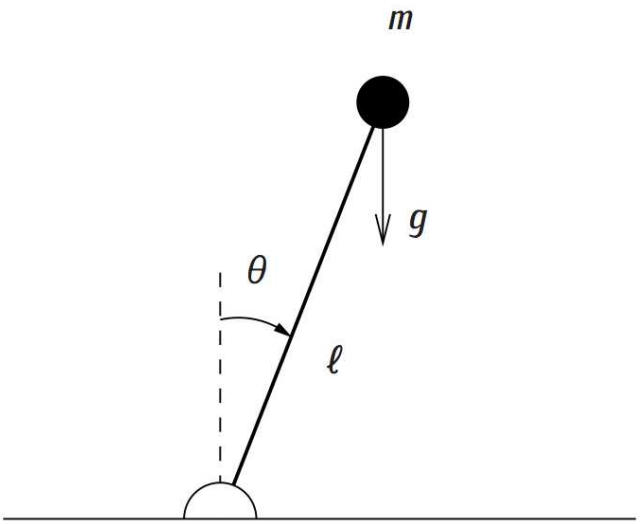


Figura 6. Ejemplo de un robot de un motor.

En un robot, las entradas típicamente corresponden a los valores angulares, velocidades y torques aplicados a cada motor, mientras que las salidas suelen representar las aceleraciones de dichos motores. Estas relaciones se obtienen de la ecuación matemática de un sistema mecatrónico, la cual se calcula a través de los métodos tradicionales mencionados anteriormente. Esta ecuación es general para cualquier sistema electromecánico, en ella intervienen fuerzas como la inercia, la Coriolis, la centrífuga, la gravitatoria, etc. Así, básicamente las RNAs se asemejan lo más posible a la ecuación mencionada, donde además contempla todas las perturbaciones externas que pueda tener, caso contrario de métodos análiticos que puede contener dichos errores. Ahora que hemos comprendido el proceso de identificación del sistema, procederemos a realizar un ejemplo práctico.

Tal como en la Figura 5, ubicamos las RNAs en paralelo al modelo real del robot (ver Figura 6). De esta manera, los datos de entrada se aplican tanto al modelo real como a las redes neuronales artificiales, donde las salidas del modelo real, obtenidas por los sensores, se restan de las salidas de las RNAs. Esta diferencia proporciona el error entre el modelo real y las RNAs, el cual se introduce en el algoritmo de entrenamiento para actualizar los pesos respectivos de las RNAs. En la Figura 7, se presenta el comportamiento del modelo del robot (el cual tiene una condición inicial de 27° en su ángulo) en relación con las redes neuronales artificiales, donde se aprecia que el modelo artificial sigue fielmente al sistema mecatrónico.

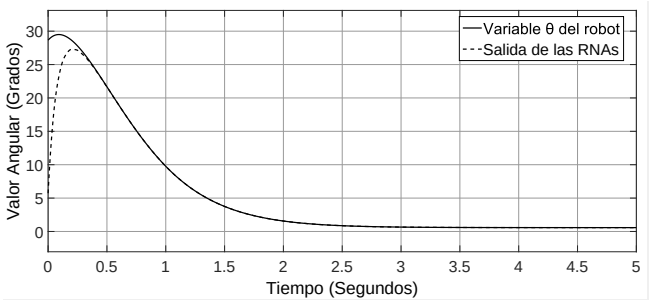


Figura 7. Comportamiento de un robot vs RNAs.

En la gráfica anterior se observa que al inicio las salidas de las RNAs no coinciden con el modelo real. Sin embargo, mediante la actualización continua, el algoritmo de aprendizaje logra ajustarse al modelo real. Esto da como resultado un sistema basado en RNAs capaz de comportarse de manera similar al modelo real. No obstante, el desafío real ocurre cuando se presentan perturbaciones en el robot, lo que puede afectar su rendimiento [9,10]. En la Figura 8, se observa el comportamiento de nuestro sistema ante perturbaciones en el sensor de posición del robot en 1s y 1.5s, donde se evidencia que, a pesar de su presencia, el algoritmo artificial replica fielmente al robot. Este escenario sería difícil de analizar mediante métodos convencionales.

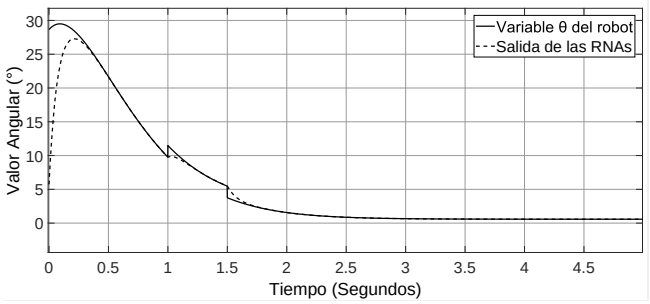


Figura 8. Comportamiento de un robot con perturbaciones vs RNAs.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

Así queda evidenciado que el uso de inteligencia artificial, específicamente hablando de redes neuronales son una solución viable para la identificación de sistemas dinámicos.

Conclusiones

En este artículo se demuestra la capacidad de las redes neuronales artificiales para la identificación de sistemas dinámicos, como es el caso de los robots. Se destaca su superioridad respecto a las técnicas tradicionales, especialmente en cuanto a adaptabilidad y su capacidad para mitigar perturbaciones externas. Mientras que en los métodos tradicionales las perturbaciones tienden a afectar el modelo, generando sistemas con errores, las redes neuronales artificiales ofrecen una mejora significativa en este aspecto.

Las perturbaciones externas pueden manifestarse de diversas formas, incluyendo variaciones en parámetros, tolerancias de materiales o desgaste de piezas a lo largo del tiempo. Además, las RNAs muestran una capacidad para adaptarse a estos cambios y mantener un rendimiento estable en la identificación y control de sistemas dinámicos. Es importante destacar que estas perturbaciones también pueden tener repercusiones en los sistemas de control a largo plazo. Si el controlador no está correctamente ajustado con respecto a las características del robot o si el modelo del sistema es incorrecto, se corre el riesgo de que el sistema falle debido a errores en los parámetros del sistema. Así, la precisa modelización y la adaptabilidad de los sistemas dinámicos son aspectos cruciales para garantizar su buen rendimiento en las aplicaciones de hoy en día. En este contexto, las RNAs se presentan como una solución para la identificación y el control de sistemas dinámicos, como es el caso de los robots. ✷

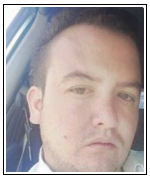
REFERENCIAS

1. Cone, E., Lambert, J. (2019). How robots change the world.
2. Kelly, R. (2003). Control de movimiento de robots manipulados. Pearson education.
3. Karur, K., Sharma, N., Dharmatti, C., Siegel, J. E. (2021). A survey of path planning algorithms for mobile robots. *Vehicles*, 3(3), 448-468.
4. Crane III, C. D., Griffis, M., Duffy, J. (2022). Screw theory and its application to spatial robot manipulators. Cambridge University Press.
5. Soori, M., Arezoo, B., Dastres, R. (2023). Artificial intelligence, machine learning and deep learning in advanced robotics, a review. *Cognitive Robotics*.
6. Simon, H. (2009). Neural networks and learning machines.
7. Alanis, A. Y., Rios-Huerta, D., Rios, J. D., Arana-Daniel, N., Lopez-Franco, C., Sanchez, E. N. (2020). High-order sliding modes based on-line training algorithm for recurrent high-order neural networks. *IFAC-PapersOnLine*, 53(2), 8187-8192.
8. Su, H., Qi, W., Yang, C., Sandoval, J., Ferrigno, G., De Momi, E. (2020). Deep neural network approach in robot tool dynamics identification for bilateral teleoperation. *IEEE Robotics and Automation Letters*, 5(2), 2943-2949.
9. Venancio, M. M., Gonçalves, R. S., da Costa Bianchi, R. A. (2020). Terrain identification for humanoid robots applying convolutional neural networks. *IEEE/ASME Transactions on Mechatronics*, 26(3), 1433-1444.
10. Liu, M., He, L., Shang, M. (2021). Dynamic neural network for bicriteria weighted control of robot manipulators. *IEEE Transactions on Neural Networks and Learning Systems*, 34(8), 4570-4583.

SOBRE LOS AUTORES



Jesús A. Medrano Hermosillo obtuvo el grado de doctor en Ciencias del Centro de Investigación y de Estudios Avanzados del IPN. Actualmente es profesor investigador en el Tecnológico Nacional de México campus Chihuahua. El Dr. Medrano es miembro del Sistema Nacional de Investigadores, donde sus intereses científicos incluyen control inteligente de sistemas y sistemas no lineales.



Abraham E. Rodríguez Mata obtuvo su doctorado en Ciencias en el Centro de Investigación y de Estudios Avanzados del IPN. En la actualidad, desempeña el cargo de profesor investigador en el Tecnológico Nacional de México, campus Chihuahua. Además, es miembro destacado del Sistema Nacional de Investigadores, centrando sus intereses científicos en sistemas no lineales, optimización y control de bioprocesos.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691



Larbi Djilali obtuvo su doctorado en Ciencias en el Centro de Investigación y Estudios Avanzados del IPN. Hoy en día, está afiliado al Tecnológico Nacional de México, en su campus Chihuahua. También es un miembro destacado del Sistema Nacional de Investigadores, con un enfoque principal en el control automático y la ingeniería de control inteligente.



Victor Alejandro Gonzalez Huitron obtuvo el grado de Doctor en Comunicaciones y Electrónica en 2017 por parte del Instituto Politécnico Nacional. En su experiencia profesional se ha desempeñado como Ingeniero de capacitación y Diseño electrónico, Ingeniero en telecomunicaciones, Analista de datos, investigador CONACYT y desarrollador de algoritmos para la industria automotriz. Actualmente es miembro del Sistema Nacional de Investigadores. Actualmente se desempeña como profesor e investigador en el Instituto Tecnológico de Querétaro. Sus intereses de investigación y desarrollo son: Visión por computadora, Inteligencia Artificial y ciencia de datos.



Oscar J. Suarez Sierra obtuvo el grado de Ingeniero Electrónico de la Universidad Pontificia Bolivariana Seccional Bucaramanga en el año 2011. En 2015, obtuvo el grado de Magister en Controles Industriales de la Universidad de Pamplona. En 2019, le fue concedido el grado de Doctor en Ciencias en Ingeniería Eléctrica por el Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV). Actualmente es Docente de la Facultad de Ingenierías de la Universidad de Pamplona en Pamplona, Colombia; Presidente del Capitulo Colombiano de Inteligencia Computacional de IEEE e Investigador Asociado del Ministerio de Ciencia Tecnología e Innovación en Colombia.

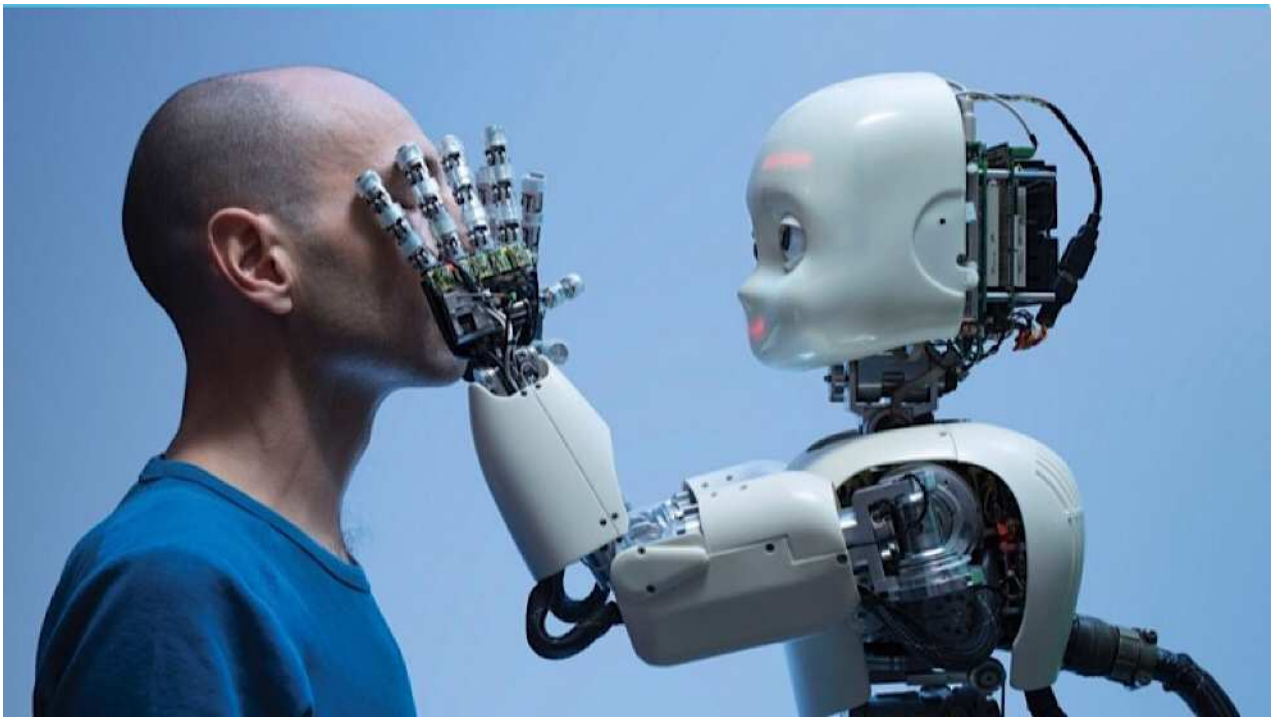


Imagen tomada de:
https://unamglobal.unam.mx/global_revista/dotan-de-inteligencia-a-los-robots-a-traves-de-redes-neuronales/.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

IA & Educación

Julieta Noguez, Lucía Barrón y Yasmín Hernández

Las Industrias 4.0 y 5.0 imponen nuevos retos a la educación

En el ámbito educativo nos enfrentamos a diversos y fascinantes retos provocados por la revolución tecnológica que está transformando a la humanidad. Las revoluciones producen cambios que tienen implicaciones en muchos aspectos de la vida. Por ejemplo, la historia nos ha mostrado el deterioro ambiental, la migración de la población o la aparición de grandes empresas como consecuencias de las diferentes etapas de la llamada Revolución Industrial. Por otra parte, también es posible observar nuevos y mejores medios de transporte y de producción, así como el incremento en la esperanza de vida con mejor alimentación y salud. Landes, citado por Chaves [1], propone que "*El término revolución industrial suele referirse al complejo de innovaciones tecnológicas que, al sustituir la habilidad humana por la maquinaria y la fuerza humana y animal por energía mecánica, provoca el paso desde la producción artesanal a la fabril, dando así lugar al nacimiento de la economía moderna*". En las diferentes etapas de la revolución industrial es posible observar los cambios generados en aspectos económicos, sociales, y culturales, por mencionar algunos. La sustitución del uso de la fuerza animal por el ferrocarril y la máquina de vapor para la locomoción lograron mejorar los alimentos, la producción, el transporte y la comunicación [1]. En la segunda etapa el uso de nuevas fuentes de energía (gas y electricidad), nuevos materiales (acero y petróleo) y nuevos sistemas de transporte, produjeron cambios en la naturaleza del proceso de producción creando la producción en masa y las líneas o cadenas de producción, y generando un importante crecimiento en la economía de varios modelos económicos [2]. La tercera etapa, a finales del siglo XX estuvo impulsada por los semiconductores, las computadoras de gran capacidad, las computadoras personales y el internet, concibiendo un mayor uso de las energías renovables, nuevas tecnologías de almacenamiento de energía, redes de distribución de energía eléctrica "inteligente" y transporte basado en vehículos eléctricos entre otros [3]. Para cada una de estas revoluciones el área de educación ha debido adaptarse y enfrentar retos con el fin de formar nuevas generaciones con los conocimientos necesarios para desempeñarse adecuadamente tanto en la sociedad como en el ámbito laboral.

Sin embargo, los retos en la educación nunca habían sido tan amplios y complejos como los que se enfrentan en esta cuarta revolución industrial o revolución digital, también llamada Industria 4.0. En el Foro Económico Mundial (WEF, por sus siglas en inglés) de 2016, Schwab [4], mencionó que esta revolución está caracterizada por el uso de un internet más ubicuo, computación en la nube, internet de las cosas, sistemas ciber-físicos, inteligencia artificial y aprendizaje máquina o *machine learning*. Aunque la denominación de esta revolución ha sido criticada tanto conceptual, histórica y socialmente por diversos autores, la realidad es que actualmente el desarrollo de la tecnología avanza de manera vertiginosa e incluso ha surgido el término de Industria 5.0.

El término *Educación 4.0* se ha relacionado con los cuatro periodos conocidos de revoluciones industriales. Éste se asocia a las metodologías, prácticas y actividades que buscan personalizar los procesos de generación de conocimiento y transferencia de información, haciéndolos más eficientes, accesibles y flexibles. Lo anterior es con el fin de desarrollar las competencias, conocimientos y habilidades que demanda la evolución y retos de la industria 4.0. Miranda et al. [5] proponen que la Educación 4.0 es el periodo actual en el que las instituciones de Educación Superior aplican nuevos métodos de aprendizaje, herramientas didácticas y de gestión innovadoras, e infraestructuras inteligentes y sostenibles complementadas principalmente por las nuevas y emergentes TIC para mejorar los procesos de generación de conocimiento y transferencia de información. La combinación de estos recursos durante los procesos de enseñanza-aprendizaje favorece la formación y el desarrollo de las competencias críticas deseables en los estudiantes actuales. Debido a lo anterior, los perfiles profesionales se hacen cada vez más específicos, destacándose la necesidad de expertos en datos, ciberseguridad o robótica para cubrir las demandas profesionales que necesita la industria y la revolución digital en las que está inmersa.

Por otra parte, las características de la Industria 5.0 implican la colaboración estrecha entre las personas y las máquinas que surgen de la inteligencia artificial. En esta revolución, se considera que las máquinas no reemplazarán a los seres humanos, sino que complementarán sus habilidades para que sean más eficientes, atendiendo las necesidades de la sociedad y personalizando la producción para evitar excedentes que pongan en peligro el ecosistema mundial.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

El informe de la Comisión Europea titulado "Industria 5.0: hacia una industria europea resiliente, centrada en el ser humano y sostenible" [6] resalta, la importancia de las personas en las fábricas inteligentes al mencionar que: "*Un requisito previo valioso para la Industria 5.0 es que la tecnología sirva a las personas en lugar de que las personas se adapten continuamente a la tecnología en constante evolución*". En relación con los beneficios que la Industria 5.0 aporta a las empresas, la Unión Europea (UE) destaca los tres más importantes: la retención de más y mejor talento humano, el ahorro de energía y el aumento de la resiliencia. Adicionalmente, se logra un beneficio global a largo plazo: competitividad y relevancia gracias a la adaptación industrial a los nuevos mercados y a un mundo en constante cambio.

El panorama de la enseñanza de la ingeniería en este ámbito se encuentra en un punto crítico que exige una reevaluación de las estrategias pedagógicas para integrar los avances tecnológicos con el desarrollo de las cualidades humanas intrínsecas. Se busca crear sinergias entre la Inteligencia Artificial (IA) y el aprendizaje centrado en el ser humano, donde los ingenieros además de ser técnicamente competentes, también deben poseer cualidades intrínsecas como la creatividad, el pensamiento crítico, la metacognición y la autorregulación útiles para dar solución a los retos de este siglo [7].

Debido a lo anterior, el futuro de la IA en la educación es abierto, amplio y complejo. A medida que la tecnología siga evolucionando, también deberán hacerlo los planteamientos para integrarla en los entornos de aprendizaje. Esta integración es vital para preparar a los estudiantes para los polifacéticos retos y oportunidades del mundo moderno y tecnológico, equipados con las habilidades interpersonales y los fundamentos éticos necesarios para navegar y dar forma a un mundo en rápida evolución.✷

REFERENCIAS

1. Chaves, P.J. (2004). Desarrollo tecnológico en la primera revolución industrial. *Revista de Historia* 17. ISSN 0213-375X

2. Bilbao, L. M. y Lanza, R. (2010). Los inicios de la Segunda Revolución Industrial, 1870- 1914 Universidad Autónoma de Madrid, ed. Historia económica, 155-196. https://repositorio.uam.es/bitstream/handle/10486/11139/55646_HistoriaEconomicaCC.pdf?sequence=1. (consultada el 20 de mayo de 2024).

3. Rifkin, J. (2018). The Third Industrial Revolution - A Radical New Sharing Economy. https://www.youtube.com/watch?v=QX3M8Ka9vUA&t=9s. (consultado el 22 de mayo de 2024).

4. Schab, K. (2017). La cuarta revolución industrial. Editorial Debate. ISBN 978-607-315-475-8.

5. Miranda, J., Navarrete, Ch., Noguez, J., Molina-Espinosa, J.M, Ramírez-Montoya, M.S., Navarro-Tuch, S.A., Bustamante-Bello, M.R, Rosas-Fernández, J.B. y Molina, A. (2021). The Core Components of Education 4.0 in Higher Education: Three Case Studies in Engineering Education. *Computers and Electrical Engineering*, (93) 107278, ISSN: 0045-7906. https://doi.org/10.1016/j.compeleceng.2021.107278.

6. Breque, M., De Nul, L. y Petridis, A. (2021). Industry 5.0 Towards a sustainable, human-centric and resilient European industry. Dirección General de Investigación e Innovación (Comisión Europea). https://op.europa.eu/es/publication-detail/-/publication/468a892a-5097-11eb-b59f-01aa75ed71a1. (consultado el 23 de mayo de 2024).

7. Balart, T. y Shryock, K.J. (2024). A framework for Integrating AI into Engineering Education, Empowering Human-Centered Approach for Industry 5.0. *Educon IEEE 2014. Proceedings of EDUCON IEEE 2024*. Kos, Grecia.

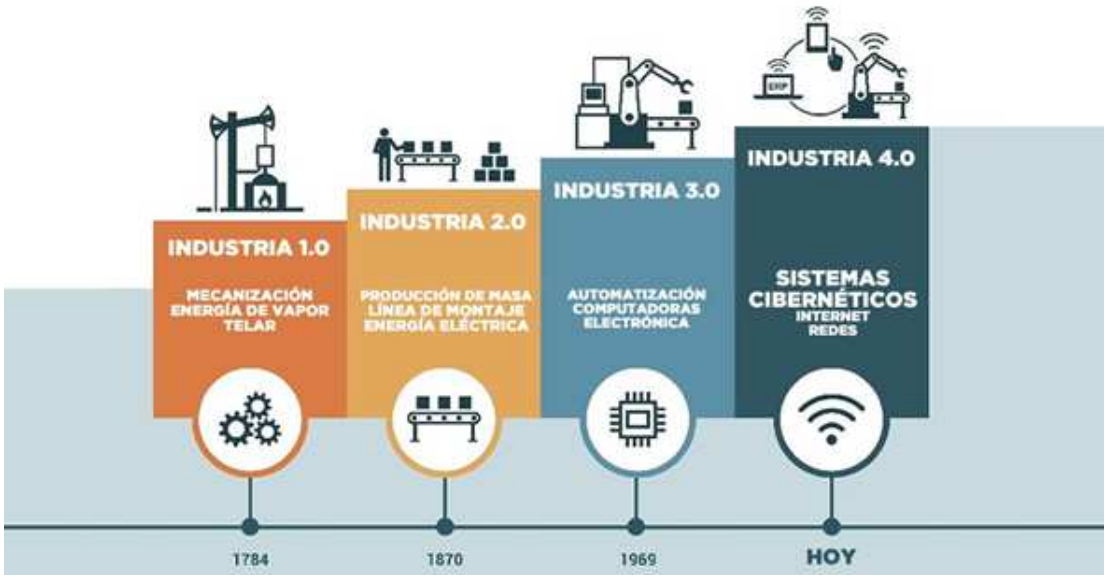


Imagen tomada de:
http://www.soymujerrural.com/mujeres-en-la-ciencia/revolucion-4-0/

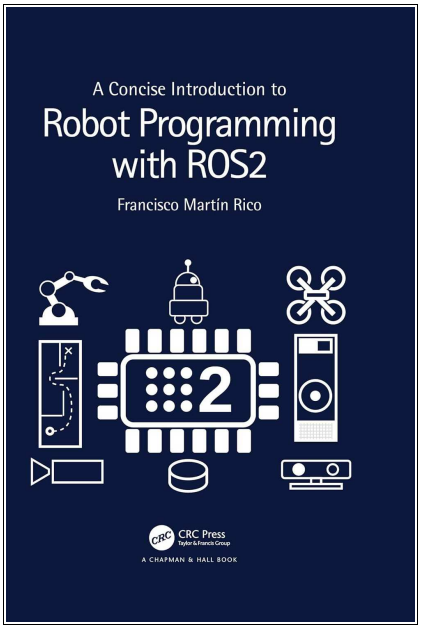
© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

Deskubriendo Konocimiento

Alejandro Guerra Hernández y **Gildardo Sánchez Ante**
deskubriendokonocimiento@komputersapiens.org

Revisión de "A Concise Introduction to Robot Programming with ROS2" *

Dr. Gildardo Sánchez Ante
Tecnológico de Monterrey
Escuela de Ingeniería y Ciencias



Portada del libro. *

Desde la popularización de los robots industriales, la programación de robots ha desempeñado un papel crucial en diversos ámbitos. Desde la manufactura hasta la exploración espacial, esta tarea ha hecho posible automatizar tareas repetitivas, mejorar la eficiencia en la producción, minimizar errores humanos y garantizar la seguridad en entornos peligrosos. Por si fuera poco, la programación también facilita la adaptabilidad de los robots a diferentes escenarios y necesidades, impulsando la innovación y el progreso tecnológico. De la optimización de procesos industriales hacia la realización de operaciones quirúrgicas precisas, el programar robots se erige como un catalizador fundamental en la evolución de la sociedad y la industria.

En un inicio cada compañía fabricante de robots diseñaba el lenguaje para programar sus robots, lo que implicaba aprender un lenguaje nuevo cada vez que se migraba de una marca de robots a otra. Esta situación motivó que en 2007 la empresa Willow Garage creara ROS (Robot Operating System), el cual es un marco de trabajo (*framework*) de código abierto cuyo propósito es el facilitar el desarrollo de software para robots. Aunque su nombre sugiere ser un sistema operativo, en realidad es un conjunto de herramientas y bibliotecas que brindan funcionalidades clave para la programación de robots, como el manejo de hardware, la comunicación entre componentes, la visualización de datos y el control de movimiento. Su diseño modular y su enfoque en la reutilización de código promueven la colaboración y la interoperabilidad en la comunidad robótica, convirtiéndolo en un estándar de facto en el desarrollo de sistemas robóticos.

ROS2 surgió en 2015 como una evolución natural de ROS, con el objetivo de abordar algunas limitaciones y desafíos que se presentaron con el tiempo en el uso de ROS original. Las compañias han ido migrando a ROS2 de manera paulatina, pero sostenida. Ante el hecho de que todos los nuevos desarrollos relacionados con ROS lo serán ya para ROS2, reflejando la importancia de migrar a esta nueva versión. Es en ese sentido que el libro *A Concise Introduction to Robot programming with ROS2*, escrito por Francisco Martín Rico, toma un rol relevante. El libro, recientemente publicado en 2022, es uno de los pocos con que se cuenta para adentrarse en los detalles más importantes de este lenguaje; está conformado por seis capítulos y un apéndice, a través de los cuales se proporciona a los lectores una introducción clara y concisa a la programación de robots utilizando ROS2. El libro se estructura de manera didáctica y práctica, abordando los conceptos fundamentales del lenguaje de una manera accesible tanto para principiantes como para profesionales.

El libro comienza abordando conceptos básicos de ROS2 que van de la instalación y configuración del entorno de desarrollo a la comprensión de la arquitectura y los componentes principales del sistema. En el segundo capítulo, se procede a crear los primeros nodos y tópicos, e incluso se lleva a cabo una prueba con

*Rico, F.M. (2022). A Concise Introduction to Robot Programming with ROS2 (1st ed.). Chapman and Hall/CRC.

© 2024 - Sociedad Mexicana de Inteligencia Artificial ISSN 2007-0691

el robot Tiago en Gazebo. Los capítulos tercero y cuarto se centran en el uso de máquinas de estados finitos para la evasión de obstáculos y en las transformaciones geométricas a través del subsistema TF, respectivamente. Los capítulos quinto y sexto se enfocan en el desarrollo de comportamientos reactivos; mientras el quinto utiliza como base la creación de un evasor de obstáculos tipo campos potenciales, el sexto lo hace mediante el uso de árboles de comportamiento (esto último permitiendo abordar problemas de SLAM). Finalmente, el apéndice contiene todos los códigos fuente de los diversos problemas analizados a lo largo del libro.

En resumen, el libro analizado es un excelente texto en el que a medida que se avanza, se exploran temas más complejos, como el desarrollo de aplicaciones robóticas completas, el manejo de sensores y actuadores, la navegación y el control de robots, entre otros. A lo largo de sus páginas el libro integra ejemplos prácticos y ejercicios que fomentan en los lectores la generación de habilidades sólidas para la programación de software de robots mediante ROS2. Aunado a lo anterior, el libro también puede incluir guías de referencia y enlaces a recursos adicionales para ampliar el conocimiento del lector en áreas específicas de interés. Sin lugar a dudas, *A Concise Introduction to Robot programming with ROS2* es una obra que enriquecería de manera excelente la biblioteca de cualquier aficionado o experto en robótica.✷



Imagen tomada del sitio:
https://www.linkedin.com/pulse/ros2-robotics-developer-course-using-python-di3dc.

© 2024 - 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



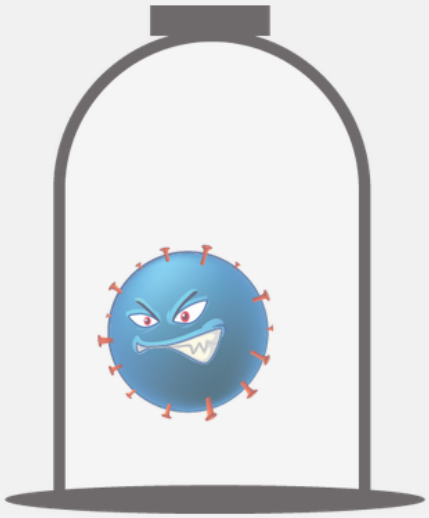
TOSE O ESTORNUDA EN LA PARTE INTERNA DE TU CODO



NO DIFUNDAS NOTICIAS FALSAS

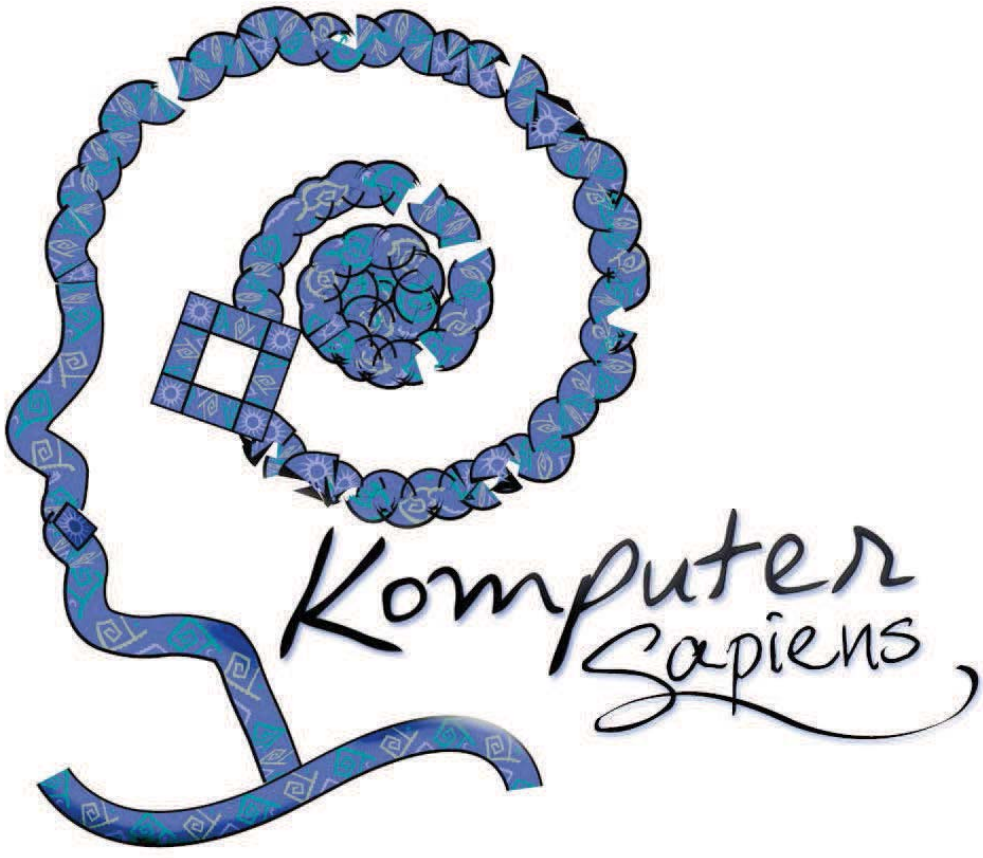
Atendiendo las normas de la Nueva Normalidad y cuidándonos unos a otros podremos salir adelante.

¡SI AISLAMOS AL VIRUS, LO VENCEREMOS!



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

Indizada en el IRMDCT de CONACYT y en Latindex



¡Publique en Komputer Sapiens!

CONAHCYT
CONSEJO NACIONAL DE HUMANIDADES CIENCIAS Y TECNOLOGÍAS

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