INTELIGENCIA ARTIFICIAL

TEORÍA Y APLICACIONES

Marco Antonio Ramos Corchado
Vianney Muñoz Jiménez
Coordinadores

Universidad Autónoma
del Estado de México

Doctor en Ciencias e Ingeniería Ambientales
Carlos Eduardo Barrera Díaz
Rector

Doctor en Ciencias Computacionales
José Raymundo Marcial Romero
Secretario de Docencia

Doctora en Farmacia y Tecnología Farmacéutica
Mariana Ortiz Reynoso
Encargada del Despacho de la Secretaría de Investigación y Estudios Avanzados

Doctor en Ciencias de la Educación
Marco Aurelio Cienfuegos Terrón
Secretario de Rectoría

Doctora en Humanidades
María de las Mercedes Portilla Luja
Secretaria de Difusión Cultural

Doctor en Ciencias del Agua
Francisco Zepeda Mondragón
Secretario de Extensión y Vinculación

Doctor en Educación
Octavio Crisóforo Bernal Ramos
Secretario de Finanzas

Maestra en Administración de Tecnologías de la Información
Jeanett Mendoza Colín
Encargada del Despacho de la Secretaría de Administración

Doctora en Ciencias Administrativas
María Esther Aurora Contreras Lara Vega
Secretaria de Planeación y Desarrollo Institucional

Doctora en Derecho
Luz María Consuelo Jaimes Legorreta
Titular de la Oficina de la Abogacía General

Maestra en Salud Animal
Trinidad Beltrán León
Secretaria Técnica de la Rectoría

Licenciada en Comunicación
Ginarely Valencia Alcántara
Directora General de Comunicación Universitaria

Maestro en Administración Pública
Juan Bernal Aguirre
Encargado del Despacho de la Dirección de Centros Universitarios y Unidades Académicas Profesionales Región A y B

Doctora en Ciencias en Ciencias Agrarias
Cristina González Pérez
Encargada del Despacho de la Secretaría de Proyectos Especiales y Proyección Universitaria

INTELIGENCIA ARTIFICIAL
TEORÍA Y APLICACIONES

Dirección de Publicaciones Universitarias
Editorial de la Universidad Autónoma del Estado de México

Doctor en Ciencias e Ingeniería Ambientales
Carlos Eduardo Barrera Díaz
Rector

Doctora en Humanidades
María de las Mercedes Portilla Luja
Secretaria de Difusión Cultural

Doctor en Administración
Jorge Eduardo Robles Alvarez
Director de Publicaciones Universitarias

INTELIGENCIA ARTIFICIAL
TEORÍA Y APLICACIONES

Marco Antonio Ramos Corchado
Vianney Muñoz Jiménez
Coordinadores

Universidad Autónoma del Estado de México
"2025, 195 años de la apertura del Instituto Literario en la ciudad de Toluca"

Primera edición, mayo 2025

INTELIGENCIA ARTIFICIAL
TEORÍA Y APLICACIONES
Marco Antonio Ramos Corchado y Vianney Muñoz Jiménez
Coordinadores

Universidad Autónoma del Estado de México
Av. Instituto Literario 100 Ote., Col. Centro
Toluca, Estado de México
C.P. 50000
Tel: 722 481 1800
http://www.uaemex.mx

Registro Nacional de Instituciones y Empresas Científicas y Tecnológicas (Reniecyt): 1800233

ISBN: 978-607-26959-6-2

Hecho en México

Director del equipo editorial: Jorge Eduardo Robles Alvarez
Coordinación editorial: Ixchel Díaz Porras
Coordinación de diseño y diseño de portada: Luis Alberto Maldonado Barraza
Corrección de estilo: Rocío Franco López
Diseño: Jarini Toledano Gil

CONTENIDO

PRESENTACIÓN 11

INTRODUCCIÓN 13
- ORIGEN DE LA INTELIGENCIA ARTIFICIAL 13
  - La inteligencia artificial hoy 14
    - *Inteligencia artificial particular* 14
    - *Inteligencia artificial general* 15
  - Estructura del libro 15

AGENTES INTELIGENTES 19
- RESUMEN 19
- DEFINICIÓN DE UN AGENTE INTELIGENTE 19
- ARQUITECTURA DE UN AGENTE INTELIGENTE 20
- TIPOS DE AGENTES INTELIGENTES 24
  - Agentes inteligentes reactivos 24
  - Agentes inteligentes cognitivos 25
- MEDIO AMBIENTE 27
- APRENDIZAJE 28
- COMPORTAMIENTO 30
- REFERENCIAS 31

TUTORES INTELIGENTES 33
- RESUMEN 33
- INTRODUCCIÓN 34
- SISTEMA CLASIFICADOR DE APRENDIZAJE (LCS) 38
- TUTOR INTELIGENTE DOTADO DE UN LCS 39
- IMPLEMENTACIÓN DE UN TUTOR INTELIGENTE EN UN CONTEXTO HISTÓRICO 42

Discusión 45
Referencias 45

Detección del rostro humano en imágenes 3D 47
Resumen 47
Introducción 48
Trabajo relacionado 49
Detección del rostro en imágenes 3D con una persona 53
Análisis de curvatura 53
Análisis por cortes 56
Análisis por segmentación 57
Resultados experimentales 59
Detección del rostro en imágenes 3D con más de una persona 60
Técnica de detección 60
Entrenamiento PCA *64*
Prueba PCA *66*
Máquina de soporte vectorial (SVM) 66
Adquisición de imágenes experimentales 68
Resultados experimentales 70
Conclusiones 71
Referencias 72

Cifrado de meta-aprendizaje en redes neuronales artificiales 77
Resumen 77
Introducción 77
Cifrado de Paillier 80
Perceptrón multicapa 82
Resultados 84
Precisión de los modelos de red 85
Tiempo de ejecución 86
Conclusiones 88
Referencias 88

APRENDIZAJE AUTOMÁTICO: TENDENCIAS
Y DESAFÍOS ACTUALES 91
Resumen 91
Introducción 91
Fundamentos del aprendizaje automático 92
Paradigmas del aprendizaje automático 93
Aprendizaje supervisado 93
Aprendizaje no supervisado 94
Aprendizaje por refuerzo 95
Aprendizaje combinado 95
Aplicaciones relevantes del aprendizaje automático 96
Desafíos y futuras direcciones 97
Tendencias actuales en aprendizaje automático 98
Conclusión 99
Referencias 100

CHATGPT Y APRENDIZAJE 103
Resumen 103
Introducción a la inteligencia artificial en el aprendizaje 104
Desafíos éticos y prácticos de la inteligencia artificial
en la educación 106
ChatGPT: un modelo de lenguaje potente para la educación 108
Aplicaciones potenciales de la IA en la educación 112
Referencias 114

HACIA UN FUNDAMENTO MATEMÁTICO
DE LA INTELIGENCIA ARTIFICIAL 115
Introducción 115
Desafíos en la inteligencia artificial 116
El panorama matemático de la inteligencia artificial 118
Redes neuronales profundas 118
Principales líneas de investigación 120
Fundamentos matemáticos de la inteligencia artificial 120

Inteligencia artificial para problemas matemáticos 121
Referencias 122

PRESENTACIÓN

La obra que se presenta surge del trabajo colaborativo del Cuerpo Académico de Sistemas Computacionales formado por investigadores de la Universidad Autónoma del Estado de México (UAEMEX), cuya área de investigación es la teoría de la inteligencia artificial (IA) y sus aplicaciones, consciente de la necesidad de formar recursos humanos con conocimientos suficientes en esta área y las necesidades propias de la sociedad para resolver problemáticas en cada una de las actividades que realizan los seres humanos.

Debido a la reciente y creciente popularidad de la IA se propone una compilación teórica aplicativa que emerge para dar solución a diferentes cuestionamientos, como: el reconocimiento facial de las personas como un sistema preventivo de seguridad; el apoyo en la adquisición de nuevo conocimiento a través del acompañamiento de tutores virtuales; el descubrimiento de características y comportamientos de individuos mediante aprendizaje automático, reconocimiento de patrones y redes neuronales; el uso del lenguaje natural para cuestionar bases de conocimientos mediante el uso del ChatGTP y, finalmente, una formalización lógica matemática que permita definir una algoritmia adecuada.

Hoy es recurrente hablar de IA aplicada a diferentes áreas, pero el uso indiscriminado de este concepto nos hace reflexionar sobre si existe una verdadera IA. En este trabajo se explica la taxonomía actual de la IA con énfasis en la inteligencia particular (IP), que tiene aplicación en múltiples dispositivos comercializados en la actualidad.

El libro está formado por ocho capítulos desarrollados por los integrantes del Cuerpo Académico de Sistemas Computacionales. En cada uno se describe la parte metodológica, teórica y aplicaciones prácticas de la IA. Las técnicas desarrolladas en cada uno de los capítulos se pueden extender a diferentes áreas y problemáticas, para usos específicos, de acuerdo con los requerimientos o necesidades que se presenten en los distintos ámbitos sociales.

El libro busca generar interés en las diferentes áreas del conocimiento en las que se incluya el concepto de IA, que puede ser extendido a otras disciplinas, pero respetando

el conocimiento de cada una de las áreas. De esta forma se invita a la comunidad científica e intelectual a proponer sistemas que utilicen la IA, para que acompañen al usuario en sus distintas actividades cotidianas, profesionales y de esparcimiento.

Integrantes del Cuerpo Académico de Sistemas Computacionales 2025, que participaron en la elaboración de este libro:

- Héctor Alejandro Montes Venegas
- José Raymundo Marcial Romero
- Marcelo Romero Huertas
- Marco Antonio Ramos Corchado
- Rosa María Valdovinos Rosas
- Vianney Muñoz Jiménez

Colaboradores invitados en la participación de los capítulos del libro, por orden de aparición:

- Félix Francisco Ramos Corchado (Cinvestav-GDL, IPN)
- Héctor Caballero Hernández (Facultad de Ingeniería, UAEMEX)
- Juan Paduano Salinas (Facultad de Ingeniería, UAEMEX)
- Graciela García Ruedas (Facultad de Ingeniería, UAEMEX)
- Javier Salas García (Facultad de Ingeniería, UAEMEX)

INTRODUCCIÓN

Origen de la inteligencia artificial

Estamos viviendo en una época en la que la tecnología es parte importante de nuestras vidas y a dicha tecnología se le llama "inteligente"; se ha vuelto común hablar de teléfonos inteligentes, televisiones inteligentes, relojes inteligentes, etc. Lo cierto es que se comenzó a hablar del concepto de inteligencia artificial (IA) en los años cincuenta, con el padre de las ciencias computacionales, Alan Turing. Después de hacer la máquina *Enigma*, formuló el concepto de la prueba de Turing (1950), con la que se puede criticar la inteligencia de una máquina: si las respuestas arrojadas por la prueba de Turing son indistinguibles de las respuestas que pueda dar un ser humano. Alan Turing imaginaba que era posible replicar las capacidades del cerebro humano como un solucionador general al utilizar las capacidades de cálculo de las computadoras.

Las teorías propuestas por Alan Turing son retomadas más tarde por John McCarthy, que después de haber trabajado en IBM comprendió que las computadoras digitales ofrecían la opción de manejar grandes cantidades de datos en periodos relativamente cortos de tiempo, utilizando las máquinas de Turing y los autómatas de Von Neumann.

Fue en 1956, durante la escuela de verano, cuando John McCarthy acuñó el término inteligencia artificial (IA), en el que se denota cómo una ciencia y/o ingeniería puede ser capaz de hacer máquinas inteligentes. John McCarthy veía la IA como una máquina que en realidad podría replicar la inteligencia humana. Uno de los dilemas con los que siempre trató McCarthy fue que los desarrollos dentro de la IA solo eran réplicas de comportamientos de los seres humanos y no verdaderas máquinas que aprendían.

La inteligencia artificial hoy

Sin duda, el avance tecnológico en los sistemas computacionales ha dado origen al desarrollo e implementación de la inteligencia artificial como la vemos hoy, son varias las aplicaciones e instrumentos de aprendizaje máquina que utilizan y procesan grandes volúmenes de datos que son almacenados en distintos repositorios, lo que se conoce como la nube (iCloud).

Asimismo, las técnicas de la inteligencia artificial son empleadas de acuerdo con la complejidad del problema por resolver. Lo cierto es que los diferentes centros de investigación, universidades y sectores industriales convergen en realizar una taxonomía de la inteligencia artificial para reconocer y afirmar que en realidad una máquina tiene la capacidad de apropiarse del conocimiento.

De acuerdo con la taxonomía hecha en el área de estudio, la inteligencia artificial se clasifica en tres grandes ramas:

- Inteligencia artificial particular
- Inteligencia artificial general
- Inteligencia artificial social

Inteligencia artificial particular

Según la clasificación de la inteligencia artificial se observa una gran utilización de inteligencias particulares. Esto quiere decir que este tipo de inteligencias son dedicadas, aprenden sobre un problema en específico (como los sistemas de recomendación), los sistemas de geolocalización, los sistemas de clasificación de datos, entre otros. Se puede decir que, la inteligencia artificial particular aborda un problema en específico sin posibilidad de cambio, esto se observa en sistemas que realizan una sola tarea.

Inteligencia artificial general

La inteligencia artificial general busca un aprendizaje universal que no solo sea la especialización para la solución específica de un problema, esto es, contar con una inteligencia capaz de adaptarse a diferentes condiciones dentro de un ambiente. La inteligencia artificial general se basa en un aprendizaje constante de acuerdo con las interacciones y acciones que se hacen dentro de un contexto, en este tipo de inteligencias emergen diferentes conceptos asociados al campo de la neurociencia, con la finalidad de poder comprender cómo es que el cerebro humano procesa la información. Para que una inteligencia artificial general pueda adaptarse a cualquier condición que se presente es necesario recurrir al concepto de cognición y memoria. La cognición se basa en cómo una inteligencia artificial aprende a partir de la percepción y acción tomada dentro de un entorno; mientras que la memoria implica almacenar solo los datos que resultan útiles en un contexto determinado.

Inteligencia artificial social

La inteligencia artificial social incorpora dentro de sus procesos de aprendizaje el factor comportamental, que le permite interactuar y cohabitar con entidades en espacios comunes. En la actualidad, la inteligencia artificial social se encuentra en desarrollo debido a las complejidades que los procesos comportamentales del cerebro humano utilizan para regular procesos como emoción, atención, toma de decisiones, entre otros.

Estructura del libro

En este libro se presentan diferentes perspectivas de cómo se emplea la IA para tratar diferentes problemáticas utilizando las técnicas y tecnologías disponibles hoy en día.

En esta "Introducción" se presenta al lector en el tema de la IA mediante un breve repaso histórico acerca de cómo se llegó a ella, su situación actual y las tres grandes ramas en que se clasifica.

En "Agentes inteligentes" se presenta el concepto de agentes, entidades autónomas capaces de exhibir comportamientos denominados inteligentes. Un agente es una entidad dotada de IA que mimetiza directamente a los seres humanos. La complejidad del agente radica en la forma en cómo se estructura una base de conocimiento inicial que le permita ir aprendiendo y adaptándose a las condiciones del medio ambiente, como lo haría un ser humano. Se presenta la arquitectura básica de un agente que permita su autonomía para cumplir objetivos de acuerdo con las necesidades específicas y la solución de problemas.

En "Tutores inteligentes" se muestra un sistema basado en agentes inteligentes que apoyan al usuario en la adquisición de nuevo conocimiento, a este tipo de sistemas se les conoce como *tutores virtuales.* Se presentan diferentes arquitecturas para la construcción de tutores virtuales, así como formas de validación de la adquisición de nuevo conocimiento. Se aborda el uso de un sistema de aprendizaje por clasificación que permita al sistema adaptarse a las necesidades del usuario, lo cual permite garantizar que a partir de un conocimiento exhibido el usuario asegure que la información es retenida y validada como aprendizaje nuevo.

En "Detección del rostro humano en imágenes 3D" se trata el concepto de la percepción, en la IA este es un elemento clave para la adquisición de información que permite a un sistema inteligente discernir y tomar decisiones dentro de un contexto. El capítulo hace uso de la percepción sustentada en 3D para la detección de rostros humanos con base en técnicas de la IA para la identificación certera, sin importar el número de individuos que aparezcan en una escena o imagen.

En "Cifrado de meta-aprendizaje en redes neuronales artificiales" se presenta una técnica de la IA ampliamente estudiada por la comunidad científica, conocida como *redes neuronales.* Una red neuronal es un sistema de aprendizaje supervisado o no supervisado que hoy en día es ampliamente utilizado para tratar grandes volúmenes de información. Se muestra el uso de las redes neuronales con la finalidad de extraer la información necesaria para la toma de decisiones, desde la premisa de no exponer los datos sensibles que pueden poner en riesgo la integridad de los usuarios.

En "Aprendizaje automático: tendencias y desafíos actuales" se aborda el concepto de cómo las máquinas aprenden y pueden generar nuevo conocimiento. Este capítulo expone las diferentes técnicas del aprendizaje máquina, así como sus

retos y las aplicaciones directas en las diferentes áreas en las que el ser humano interviene y requiere ser asistido.

En "ChatGPT y aprendizaje" se retoma la propuesta de OpenAI con su desarrollo ChatGPT para su análisis y valoración dentro del sector educativo. En la actualidad, ChatGPT está siendo utilizado en diferentes áreas del conocimiento, en específico, en la adquisición de nuevo conocimiento por parte de los estudiantes; esta tecnología ofrece grandes retos tanto en lo técnico como en lo social, sobre todo, en cuanto al uso correcto de esta tecnología y el sentido ético de los usuarios.

Por último, en "Hacia un fundamento matemático de la inteligencia artificial" se trata el aspecto formal en la construcción de entidades inteligentes y autónomas, el lenguaje matemático ha sido una forma universal de expresar fenómenos físicos y naturales para estructurar funciones y conceptos a partir de un lenguaje natural. En este capítulo se hace una extensión del formalismo simbólico-matemático aplicado a la inteligencia artificial.

AGENTES INTELIGENTES

Marco Antonio Ramos-Corchado
Facultad de Ingeniería, UAEMEX

Vianney Muñoz-Jiménez
Facultad de Ingeniería, UAEMEX

Félix F. Ramos
Facultad de Ingeniería, UAEMEX

Resumen

La IA retoma el concepto de agente inteligente como una entidad completamente autónoma, capaz de realizar una tarea o resolver un problema de forma inteligente dentro de su medio ambiente, utilizando mecanismos de percepción, tal como lo hacen los seres humanos.

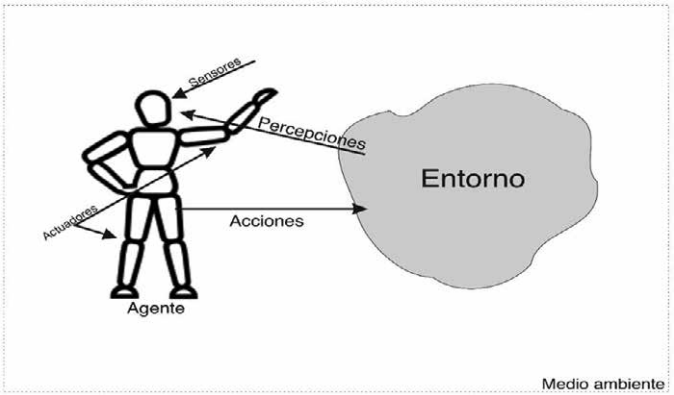
Definición de un agente inteligente

La definición de un agente inteligente se concibe como un sistema perceptivo, que permite la interpretación y procesamiento de la información que recibe de su entorno, para actuar en consecuencia de acuerdo con los datos que recolecta y procesa. El comportamiento de un agente inteligente es reflexivo, esto significa que es capaz de actuar de forma independiente al exhibir control sobre su morfología inicial sin olvidar en todo momento su objetivo principal (Dellaert y Beer, 1996; Gravina, Liapis y Yannakakis, 2018).

Arquitectura de un agente inteligente

La figura 2.1 ilustra la arquitectura básica de un agente inteligente conformado por el medio ambiente, un entorno local, un conjunto de sensores, un conjunto de actuadores y una base de conocimiento inicial. El medio ambiente representa el ecosistema, mientras que el entorno local engloba la percepción del agente inteligente, los sensores le permiten observar su medio ambiente y la base de conocimientos iniciales le permite construir posibles acciones para lograr sus objetivos a través de un conjunto de actuadores.

Figura 2.1. Arquitectura básica de un agente inteligente



Fuente: elaboración propia.

Sin importar el problema que se desee tratar empleando un sistema de agentes inteligentes, su arquitectura debe ser flexible, es decir, que muestren la capacidad de reactividad, proactividad y sociabilidad. La reactividad permite que los agentes inteligentes sean altamente reactivos a las condiciones del entorno. Donde un sistema catalogado como reactivo es aquel que conserva una interacción continua con su medio ambiente y reacciona a los cambios que se producen en él, con un tiempo conveniente para que la respuesta sea útil.

La proactividad permite que el agente inteligente tome acciones de tipo racional en condiciones de incertidumbre para evitar estados de bloqueo, para ello, es necesario

contar con sistemas reactivos al medio ambiente, de tal forma, que toda acción genera una reacción; es decir, todo estímulo produce una serie de respuestas (estímulo → reglas de respuesta). La proactividad implica generar y tratar de alcanzar metas u objetivos específicos que reconozcan oportunidades dentro del medio ambiente para ser alcanzados en periodos de tiempo convenientes.

La habilidad social en los agentes inteligentes consiste en la capacidad de interactuar y negociar con otros agentes inteligentes que se encuentran en el medio ambiente, utilizando algún tipo de lenguaje de comunicación entre ellos, para abrir la posibilidad de cooperar con otros agentes inteligentes con la finalidad de resolver y cumplir con los objetivos establecidos en su base de conocimientos iniciales.

Los agentes inteligentes cuentan con una base de conocimiento inicial conformada por: su entorno (*E*), las posibles acciones a realizar (*Ac*) y una función de ejecución (*r*). El entorno *E* se define por un conjunto de estados finitos, en el que se albergan los agentes inteligentes que participan en la solución de objetivos, definida por:

$$E = \{e_0, e_1, e_2, \ldots, e_u\}$$

Los agentes inteligentes tienen la posibilidad de generar o construir un repertorio de posibles acciones *Ac* de acuerdo con la percepción que reciben del entorno, modificando su estado inicial por:

$$Ac = \{\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_{u-1}\}$$

Los agentes inteligentes realizan las acciones dentro del *E* a través de la construcción de una lista que procesa la función de ejecución *r*, que corresponde a una secuencia de estados y acciones que pueden ser realizados dentro del *E* para el logro de sus objetivos. La lista de funciones de ejecución *r* se define como:

$$r = e_0 \overset{\alpha_0}{\rightarrow} e_1 \overset{\alpha_1}{\rightarrow} e_2 \overset{\alpha_2}{\rightarrow} e_3 \overset{\alpha_3}{\rightarrow} \ldots e_{u-1} \overset{\alpha_{u-1}}{\longrightarrow} e_u$$

Los agentes inteligentes no cuentan con el contexto general de dónde se encuentran situados, esto es porque su percepción se debe basar en el estado inicial e_0, para ello, debe construir modelos abstractos a partir de la percepción local. Donde, *R* es el

conjunto de todas las sucesiones finitas posibles de *E* y *Ac.* Se construyen sucesiones *RAc* que terminan en una acción y aquellas que permiten terminar en un estado del entorno, *RE.*

Las acciones realizadas por los agentes inteligentes generan cambios directos e indirectos dentro del medio ambiente, a estos cambios se les denomina comportamiento del entorno τ, considerando que los entornos son dependientes de un histórico $\wp$ y no deterministas. τ se expresa de la siguiente manera:

$$\tau : R^{Ac} \rightarrow \wp(E)$$

Cuando no existen posibles estados sucesores $\tau(r) = \emptyset$, en este caso, el sistema ha terminado de ejecutarse y se validan los resultados que se obtuvieron. Formalmente, el medio ambiente *Env* es una tripleta donde *E* es un conjunto de estados del entorno, $e_0 \in E$ que corresponde al estado inicial y τ se define como la función de comportamiento que modifica el estado inicial. El medio ambiente se define por:

$$Env = \langle E, e_0, \tau \rangle$$

Los agentes inteligentes *Ag* tienen un comportamiento respecto al medio ambiente *Env,* es decir, pueden tomar decisiones sobre qué acción deben realizar en función del historial de las percepciones del *Env* hasta la fecha. Contar con un histórico le permite al *Ag* desencadenar una serie de acciones sin necesidad de recalcular una acción cada vez que se cambia de estado. El conjunto de todos los agentes inteligentes (*Ag*) se define por:

$$Ag : R^E \rightarrow Ac$$

Este tipo de sistema se compone por agentes inteligentes *Ag* y el medio ambiente *Env*, en este, es posible hacer una serie de acciones por parte del *Ag* en un tiempo adecuado para el logro de objetivos *R(Ag, Env)*. Donde las acciones son finitas, por ejemplo: $(e_0, \alpha_0, e_1, \alpha_1, e_2, \alpha_2, \ldots)$, lo que representa una serie de estados y acciones a realizar por *Ag* dentro del $Env = \langle E, e_0, \tau \rangle$. Por ejemplo:

1: e_0 es el estado inicial del *Env*
2: $\alpha_0 = Ag(e_0)$
3: Para $u > 0$

$$e_u \in \tau((e_0,\alpha_0,\ldots,e_{u-1},\alpha_{u-1})) \qquad \alpha_u = Ag((e_0,\alpha_0,\ldots,e_{u-1},\alpha_{u-1}))$$

Los agentes inteligentes tienen la capacidad de percibir el *Env*, por lo que cuentan con una función que les permite observar y definir un proceso de decisiones, la función de percepción se define como:

$$ver : E \rightarrow Perc$$

Es decir que se asignan los estados del entorno a las percepciones y las acciones ahora pasan a ser una función:

$$f(Ac) : Perc* \rightarrow A$$

Con esto se generan secuencias de percepciones a acciones con lo que los *Ag* muestran comportamientos naturales que les permiten alcanzar los objetivos definidos.

De forma general, el control interno de los *Ag* se observa como una secuencia de pasos de la siguiente forma:

1. El *Ag* comienza en algún estado interno inicial e_0
2. Observa el estado de su entorno *E,* y genera una percepción ver(*E*)
3. Entonces, el estado interno del *Ag* se actualiza a través de una función siguiente $sig(e_0, ver(E))$
4. La acción seleccionada por el *Ag* es $Ac \rightarrow sig(e_0, ver(E))$
5. Regresa a paso 2 hasta que alcanza el objetivo

Como se ha descrito con anterioridad, un agente inteligente tiene la capacidad de ser consciente de su medio ambiente e interactuar sobre él mediante la generación de diversas acciones; sin embargo, es imprescindible que la toma de decisiones sea equilibrada de acuerdo con el costo-beneficio para el logro de objetivos y tareas.

Tipos de agentes inteligentes

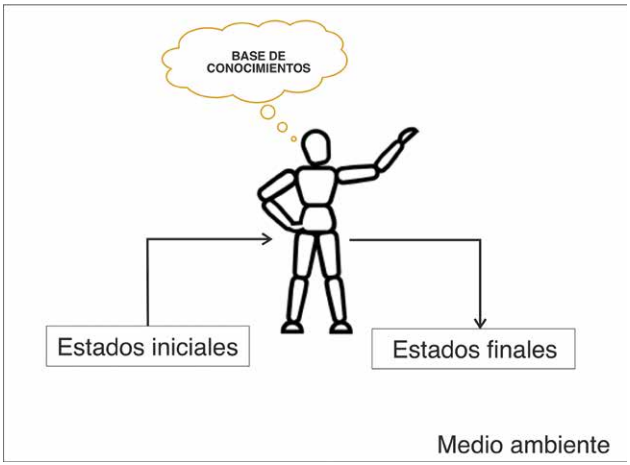
Hoy en día existen diferentes tipos de agentes inteligentes construidos a partir de una necesidad específica, en esta sección trataremos dos de los principales tipos de agentes inteligentes: reactivos y cognitivos.

Agentes inteligentes reactivos

Los agentes inteligentes reactivos son los que se implementan de manera más básica, debido a que no requieren una representación global del medio ambiente. Las acciones que ejecutan este tipo de agentes inteligentes se basan en reaccionar directamente a las percepciones del medio ambiente. El logro de objetivos se basa en la interacción de varios agentes inteligentes, las colonias de hormigas y termitas son ejemplos claros de agentes inteligentes reactivos (Adami, Brown y Kellogg, 1994).

La figura 2.2 muestra la arquitectura de un agente inteligente reactivo. La emergencia de la IA en este tipo de agentes se basa en la cantidad de agentes inteligentes que participan en la realización de tareas y el logro de objetivos dentro del medio ambiente. Es común confundir a los agentes inteligentes reactivos con los sistemas expertos, la principal diferencia es su organización. En un sistema experto se utiliza una base de conocimiento tipo tabla, donde se encuentra la totalidad de información del medio ambiente y cualquier modificación no validada por el experto puede desestabilizar el sistema. Mientras que, en un sistema multiagentes, cada agente inteligente posee una parte del medio ambiente que puede compartir con otros agentes inteligentes sin poner en riesgo la estabilidad del sistema y lograr sus objetivos.

Figura 2.2. Arquitectura de un agente inteligente reactivo



Fuente: elaboración propia.

Agentes inteligentes cognitivos

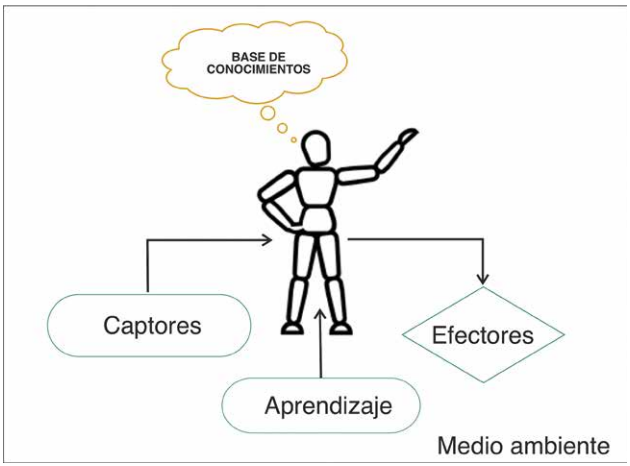
Los agentes inteligentes cognitivos, en general, cuentan con una representación global de su medio ambiente y cada agente inteligente sabe realizar una tarea, además de tener una base de conocimientos inicial. Este tipo de agentes inteligentes son intencionales debido a que tienen objetivos particulares. Una de las características importantes de estos agentes inteligentes es la capacidad de aprender a utilizar la representación del medio ambiente para adaptarse. Los agentes inteligentes cognitivos son interesantes debido a su capacidad para poder representar y adquirir conocimiento de diferentes formas (Lessin, Fusell y Miikkulainen, 2014).

La figura 2.3 muestra la arquitectura básica de un agente inteligente cognitivo. A estos agentes inteligentes se les puede dotar de algoritmos que les permitan aprender de las acciones que se realizan en el medio ambiente y utilizarlas para el logro de objetivos individuales.

Los agentes inteligentes cognitivos deben ser completamente autónomos, esto es, deben realizar sus acciones y tomar sus decisiones de forma individual. La capacidad de generar sus propios estados internos en función de la percepción

provoca el surgimiento de comportamientos particulares en cada uno de los agentes inteligentes. Un ejemplo son los agentes inteligentes que se encuentran en la web, estos proporcionan ayuda a los usuarios de acuerdo con sus necesidades particulares, como es la clasificación de la información de acuerdo con sus intereses particulares.

Figura 2.3. Arquitectura de un agente cognitivo



Fuente: elaboración propia.

En general, los agentes inteligentes cognitivos deben cubrir las siguientes características:

- Son entidades que deberán cumplir con objetivos en ambientes dinámicos y complejos.
- Muestran una autonomía en la ejecución de acciones y toma de decisiones.
- Son capaces de adaptarse a las condiciones del entorno y el medio ambiente.
- La adaptación requiere atender y resolver problemas complejos, pero también tomar en cuenta los cambios dinámicos del medio ambiente.

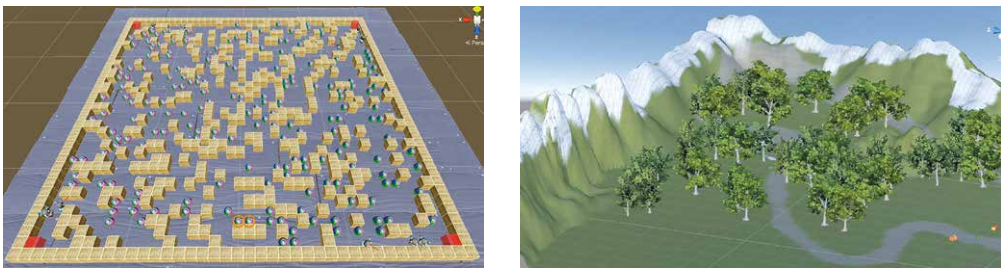
Medio ambiente

Las arquitecturas de agentes inteligentes utilizan el medio ambiente, esto es, en donde cada agente debe aprender a desempeñar un rol. No podemos construir agentes inteligentes sin contar con el medio ambiente en el que estarán embebidos y las relaciones existentes entre ellos; es decir que el medio ambiente es un espacio en donde se encuentran objetos pasivos o activos con reglas de interacción (Pilat y Jacob, 2008).

Los sistemas basados en agentes inteligentes son embebidos en ambientes dinámicos y discretos. En un ambiente discreto se conoce la totalidad de componentes que conforman el espacio, así como las reglas de interacción. Dentro de los ambientes dinámicos se incorpora la noción de tiempo, esto hace que el medio ambiente cambie tanto en estructura como en reglas de interacción.

En la figura 2.4 *a)* se ilustra un ejemplo de un medio ambiente discreto, y *b)* un ejemplo de un medio ambiente dinámico, es aquí donde los agentes inteligentes deben interactuar para cumplir con una tarea u objetivo. Estos tipos de medio ambientes son considerados visuales, porque permiten observar los comportamientos de los agentes inteligentes. Sin embargo, existen otros tipos de medio ambientes en los que los agentes inteligentes basan sus objetivos para ofertar los servicios demandados por los usuarios.

Figura 2.4. Tipos de medio ambientes virtuales: *a)* discreto y *b)* dinámico



a) *b)*

Fuente: elaboración propia.

La interacción del medio ambiente con los agentes inteligentes es directa, esto se debe a que en el medio ambiente se encuentran recursos que deberán utilizar los agentes inteligentes para alcanzar sus objetivos o tareas. Los recursos pueden ser variados; por ejemplo, un valor energético, una herramienta, un espacio de memoria, entre otros. Es normal que los objetivos de los agentes inteligentes estén vinculados a los recursos disponibles en el medio ambiente, en consecuencia, el medio ambiente debe ser capaz de modificarse, local o globalmente, de acuerdo con el consumo de los recursos. Por ejemplo, si un agente inteligente encuentra un valor energético en cierto sector, este existirá hasta que se agote y, en consecuencia, desaparecerá, y el medio ambiente se modificará de forma global hasta que existan de nuevo las condiciones para que reaparezca este valor energético. Es tarea de los agentes inteligentes utilizar de forma correcta los recursos con los que cuenta el medio ambiente para el logro de sus objetivos y su persistencia.

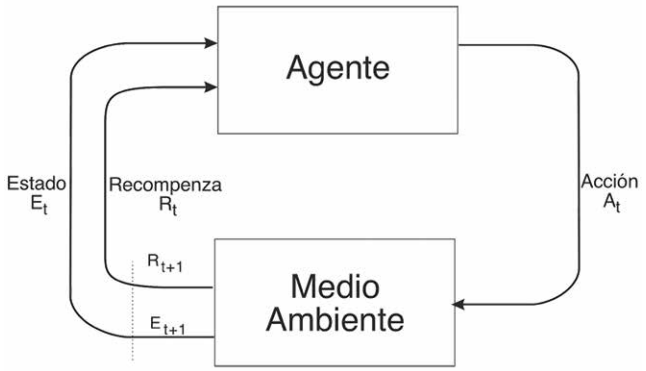
Las interacciones son reacciones recíprocas entre dos o más sistemas, esto hace que se requiera de reglas de interacción, sobre todo cuando son varias las entidades que cohabitan en el medio ambiente. Actualmente, podemos identificar tres tipos de interacciones: *1)* el medio ambiente con sus agentes inteligentes, *2)* la interacción entre agentes inteligentes, y *3)* el usuario con los agentes inteligentes. La razón de definir reglas de interacción está centrada en la parte de aprendizaje de los agentes inteligentes.

Aprendizaje

El aprendizaje en los agentes inteligentes se basa en la experiencia adquirida mediante las interacciones realizadas dentro del medio ambiente para alcanzar de forma satisfactoria el logro de sus objetivos, por ejemplo, el aprendizaje rápido (*quick learning*) que se ilustra en la figura 2.5. En la actualidad la IA propone mecanismos de aprendizaje, todos concebidos con un propósito particular, que se pueden embeber en los agentes inteligentes para hacerlos eficientes. En general, un método de aprendizaje consiste en evaluar todas las posibles acciones que se pueden realizar y seleccionar la mejor entre ellas.

Como se observa en la figura 2.5, se cuenta con un mecanismo de recompensa que permite ir mejorando las acciones dentro del medio ambiente, para alcanzar rápidamente los objetivos de los agentes inteligentes. En efecto, el aprendizaje rápido está basado en prueba y error, pero es la reactividad que poseen los agentes inteligentes la que les permite que el proceso de aprendizaje converja con rapidez hacia la solución de problemas y logros de objetivos. La simplicidad del aprendizaje rápido guía las acciones del agente inteligente para maximizar el valor esperado de las recompensas futuras. El agente inteligente aprende a estimar el valor de cada acción posible en un estado específico. Sin embargo, como el aprendizaje rápido se basa en las interacciones percibidas del medio ambiente, por los valores asociados a una función de utilidad que se almacenan en una tabla, el espacio de almacenamiento de la tabla decrece rápidamente. Este decrecimiento hace que la toma de decisiones del agente inteligente requiera cálculos adicionales debido al aprendizaje para evitar acciones no correctas durante la toma de decisiones.

Figura 2.5. Aprendizaje rápido (*quick learning*)



Fuente: elaboración propia con base en Hattori, Hagiwara y Nakagawa, 1994.

El aprendizaje es parte fundamental de cualquier sistema que se llame inteligente, la IA lo denomina aprendizaje máquina y busca las mejores técnicas para implementar este proceso, que permita a los agentes inteligentes exhibir diferentes comportamientos.

Comportamiento

La generación de comportamientos autónomos no es una tarea sencilla en un sistema de agentes inteligentes, dado que el comportamiento se puede dar de forma consciente y, en algunos casos, de forma inconsciente, también puede ser voluntario o involuntario. El comportamiento consiste en el conjunto de todas las respuestas que puede ofrecer un agente inteligente en su relación con el medio ambiente. El comportamiento es la forma de actuar de cada agente inteligente con el resto de los agentes inteligentes que comparten el medio ambiente. Un sistema de agentes inteligentes debe tener la cualidad de poder interactuar con el resto de los agentes inteligentes de forma comunicativa, cooperativa y coordinada, cuidando de no distraerse de sus propios objetivos, a esto se le conoce como *socializar*.

De igual forma que con los seres humanos, los agentes inteligentes se comportan de diferente manera, así que en un medio ambiente se debe buscar el equilibrio para no agotar los recursos disponibles que arriesguen la persistencia de los agentes inteligentes dentro del medio ambiente, con lo que se provoque la falla en el logro de sus objetivos. A partir del comportamiento del agente inteligente se podrán definir sus acciones ante ciertos estímulos provenientes del medio ambiente.

El comportamiento del agente inteligente se define como un conjunto de todos los actos exhibidos y observados en los seres vivos y los seres humanos, así como en el contexto del entorno en el que se encuentra embebido. Los comportamientos más utilizados hasta el momento son: elitista, egoísta y grupal. El *comportamiento elitista* se basa en interactuar solo con agentes inteligentes que aporten información al problema que se debe resolver. El *comportamiento egoísta* se centra en objetivos específicos, sin considerar las afectaciones causadas por otros agentes inteligentes y del medio ambiente. El *comportamiento grupal* involucra a un conjunto de agentes inteligentes con diferentes objetivos; sin embargo, la colaboración entre ellos permite ir alcanzado cada uno de sus objetivos, generalmente, este tipo de comportamientos se implementan cuando el medio ambiente es dinámico.

En la actualidad se estudian comportamientos más complejos que permitan resolver tareas y problemas de difícil solución, en los que estén implicados una gran cantidad de agentes inteligentes dotados de diferentes tipos de emociones (Amthor *et al.,* 2020). Por ejemplo, en un sistema de entrenamiento, si los agentes inteligentes

perciben resistencia en los usuarios, estos pueden modificar su comportamiento para que el usuario o usuarios no se sientan invadidos al momento de ejecutar una acción, esto se puede ver en los sistemas conocidos como guías o tutores virtuales, que se abordarán en el siguiente capítulo.

Referencias

Adami, C., Brown, C. y Kellogg, W. (1994). Evolutionary learning in the 2D artificial life system "Avida". *Artificial life IV.* Vol. 1194. MIT Press Cambridge, pp. 377-381. DOI: https://doi.org/10.48550/arXiv.adap-org/9405003

Amthor, F. R. *et al.* (2020). *Edición Essentials of Modern Neuroscience.* McGraw-Hill.

Dellaert, F. y Beer, R. D. (1996). A developmental model for the evolution of complete autonomous agents. *Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior.* MIT Press Cambridge, pp. 393-401. DOI: https://doi.org/10.7551/mitpress/3118.003.0048

Gravina, D., Liapis, A. y Yannakakis, G. N. (2018). Fusing novelty and surprise for evolving robot morphologies. *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 93-100. DOI: https://dl.acm.org/doi/10.1145/3205455.3205503

Hattori, M., Hagiwara, M. y Nakagawa, M. (1994). Quick learning for bidirectional associative memory. *IEICE TRANSACTIONS on Information and Systems, 77*(4), pp. 385-392.

Lessin, D., Fussell, D. y Miikkulainen, R. (2014). Trading control intelligence for physical intelligence: Muscle drives in evolved virtual creatures. *Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation.* ACM, pp. 705-712. DOI: https://dl.acm.org/doi/10.1145/2576768.2598290

Pilat, M. L. y Jacob, C. (2008). Creature Academy: A system for virtual creature evolution. *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence),* pp. 3289-3297.

TUTORES INTELIGENTES

Vianney Muñoz-Jiménez
Facultad de Ingeniería, UAEMEX

Marco Antonio Ramos-Corchado
Facultad de Ingeniería, UAEMEX

Héctor Caballero Hernández
Facultad de Ingeniería, UAEMEX

Resumen

En la actualidad, la adquisición de nuevo conocimiento o habilidades para la realización de trabajos o tareas resulta imprescindible para los seres humanos. La forma en que se adquieren estos conocimientos es acudiendo a los centros educativos o mediante videos tutoriales de acceso directo en la web. En los dos casos, la forma de evaluar la adquisición del conocimiento es mediante una serie de preguntas, formularios o prácticas acordes con el contexto de la formación. Sin embargo, si el usuario falla o se equivoca en las respuestas proporcionadas, es obligado a repetir completamente el entrenamiento para ser evaluado de nuevo. En muchos de los casos, los usuarios que no obtienen la calificación suficiente abandonan su formación, lo que les acarrea problemas sociales, económicos y psicológicos. Existen varias propuestas tecnológicas para minimizar este tipo de problemas, no obstante, los sistemas con los que se cuenta son estáticos y solo provocan desinterés de los usuarios, al momento de adquirir su conocimiento. Por eso en este capítulo se discute la importancia de diseñar un tutor inteligente desde el paradigma de agentes virtuales. El objetivo del tutor inteligente es reconocer las habilidades y debilidades del usuario, al momento de adquirir el conocimiento, con la finalidad de adecuar los contenidos de la unidad de aprendizaje para que la evaluación del usuario arroje un resultado positivo.

Introducción

Son muchas las razones que han hecho que los seres humanos tengan que adaptarse a las nuevas condiciones de trabajo y colaboración en las grandes ciudades, esto implica adquirir habilidades rápidamente para desempeñar alguna actividad o trabajo. Universidades, entidades certificadoras y escuelas de entrenamiento ofrecen una gran variedad de cursos especializados que permiten preparar a los usuarios en diferentes áreas del sector social, sin embargo, las necesidades económicas de los seres humanos hacen que los tiempos de entrenamiento sean realmente cortos, lo que no permite que se aproveche al máximo dicha capacitación. Para atender esto, el gobierno y los centros educativos generan contenidos directamente en una página web que es de fácil acceso para cualquier tipo de usuario.

En este tipo de adquisición de conocimientos son varios los retos que se deben afrontar, aunque la mayoría de ellos se basan en el concepto de brindar educación a distancia, es decir, existe un profesor o tutor que va guiando el aprendizaje de forma asincrónica, este se ve rápidamente rebasado por la cantidad de usuarios que utilizan el material de aprendizaje para competir por un puesto laboral. Otro reto que se debe de resolver es cómo se valida el conocimiento que se está adquiriendo por parte del usuario, por lo general, este se basa en el cuestionamiento directo a través de un test, cuestionario o examen para el usuario y se pondera el puntaje obtenido. En caso de que la calificación obtenida no sea suficiente, el usuario deberá repetir el curso o la capacitación, es aquí donde el usuario pierde interés y surge una serie de problemas emocionales (Esteve Zarazaga *et al.,* 2001). La educación a distancia, hoy conocida como educación virtual, se ve beneficiada con los avances tecnológicos, el paradigma es el mismo pero con algunos componentes automatizados y una evaluación continua, que mejora las posibilidades de concluir la formación.

La validación continua se sigue haciendo mediante cuestionarios y con la obtención de puntajes suficientes, pero ello no garantiza que en realidad el conocimiento haya sido adquirido correctamente por el usuario y que este pueda ser utilizado en consecuencia. Para subsanar esta dificultad, la IA propone tutores virtuales que acompañen al usuario durante su curso, capacitación o certificación. Los tutores virtuales son entidades autónomas que se encuentran embebidas dentro de un ambiente virtual que ha sido desarrollado expresamente para los sistemas de educación virtual. Los primeros

tutores virtuales se concibieron como motivadores que alentaban al usuario a concluir su formación; hoy en día, los tutores virtuales son utilizados para generar los cursos virtuales de forma automática con apoyo de la IA. Si bien la IA mejora la interacción con los usuarios y genera contenidos de acuerdo con las necesidades de los diferentes sectores sociales, es necesario que los usuarios tengan la posibilidad de adquirir el conocimiento de acuerdo con sus propias habilidades y que los contenidos sean adecuados (*ad hoc*) a estas habilidades. El logro de esta tarea no es sencillo, pero las técnicas de la IA posibilitan la creación de estos nuevos sistemas. En este capítulo se aborda la pertinencia de la construcción de un tutor virtual que funja como tutor en la adquisición de conocimiento de los usuarios en los ambientes virtuales.

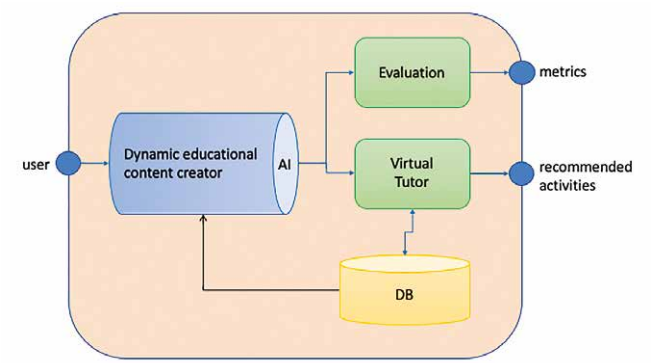
La novedad de la propuesta radica en que el tutor virtual pueda detectar las necesidades específicas de los usuarios y pueda apoyarlos en aquellos conceptos o tareas que deban ser reforzados, para ello, se hace uso de las técnicas de aprendizaje propuestas por la IA, en específico, los llamados sistemas clasificadores de aprendizaje (LCS, por sus siglas en inglés).

Tutores inteligentes

Los tutores inteligentes son agentes que utilizan las tecnologías de la información y las comunicaciones (TIC) para presentar contenidos educativos que los usuarios deben aprender (Nedungadi y Remya, 2015). En la mayoría de los sistemas educativos se hacen algunas pruebas de desempeño, sobre todo psicométricas, para que el sistema determine de forma correcta las características que deben exhibir los contenidos temáticos y/o las actividades que se deben ejecutar de acuerdo con las características que reportan los usuarios. Los programas o algoritmos que utilizan los tutores inteligentes tienen la tarea de emplear la información que obtienen acerca del desempeño de los usuarios en el sistema, para poder generar las siguientes actividades que debe hacer el usuario. Los tutores inteligentes conforman la base del paradigma "educación personalizada", en que el contenido que se genera y la velocidad de presentación de estos deben adaptarse a cada usuario. En las pruebas piloto que se han hecho se muestra que la educación personalizada (que en este caso puede ser a través del tutor inteligente) minimiza la tasa de deserción escolar.

La arquitectura básica de un sistema para generar contenidos educativos se compone de diferentes módulos, en donde destaca el tutor virtual, su función principal se basa en evaluar constantemente al estudiante y proponer actividades para reforzar el aprendizaje, en la figura 3.1 se ilustran los módulos que componen un generador inteligente de contenidos educativos.

Figura 3.1. Arquitectura básica de un generador de contenidos educativos

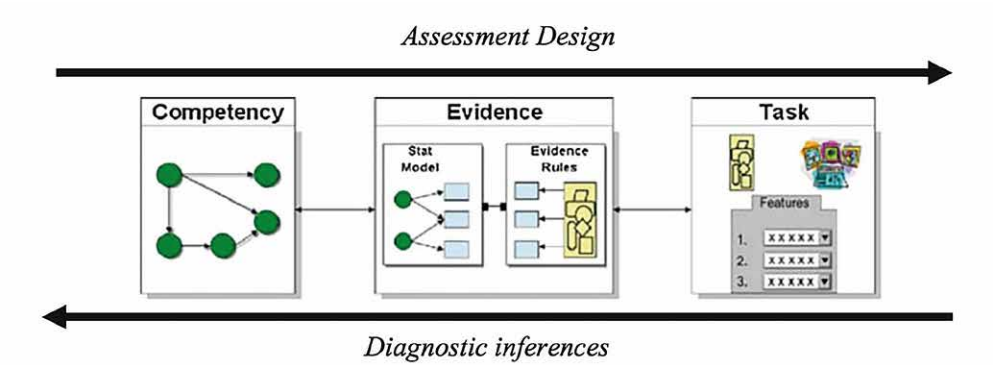


Fuente: Nedungadi y Remya, 2015.

Es fácil observar que el uso de las técnicas de IA se centra en la creación de contenido, para que el usuario pueda tener acceso rápidamente al conocimiento, mientras que el tutor virtual se centra en ver qué es lo que hace el usuario para proponer actividades de refuerzo y medir qué tanto se ha adquirido el conocimiento.

En 2017, Valerie (Shute, Rahimi y Emihovich, 2017) propuso otras arquitecturas para asegurar que el usuario adquiera las habilidades y el conocimiento, en las que se presenta un sistema basado en reglas que permite dar un seguimiento al usuario de acuerdo con las respuestas que proporciona. El sistema es capaz de generar nuevos contenidos para reforzar aquellos conocimientos en los que el usuario presenta dificultades. En la figura 3.2 se muestra la arquitectura propuesta por Valerie.

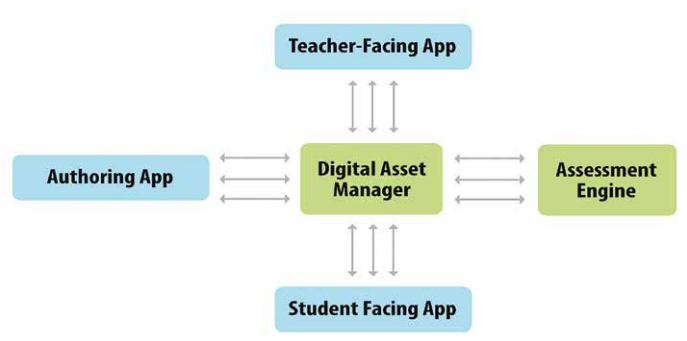
Figura 3.2. Arquitectura para validar competencias propuesta por Valerie



Fuente: Shute, Rahimi y Emihovich, 2017.

Otras aproximaciones en la generación de sistemas de entrenamiento o educativos se centran en la infraestructura con la que se cuenta, con la finalidad de que el usuario pueda tener acceso a los contenidos de su interés, tal es el caso de las plataformas digitales de aprendizaje (Richards, 2017). Este es un punto importante por resolver, ya que, no podemos pensar en la construcción de tutores virtuales sin considerar en donde estarán inmersos. En la figura 3 se muestran los componentes básicos de infraestructura que deben contener las plataformas de aprendizaje.

Figura 3.3. Arquitectura para plataformas digitales de aprendizaje



Fuente: Richards, 2017.

La IA se encuentra embebida en los sistemas educativos para el aprendizaje; sin embargo, resulta evidente que se extienda a la interacción directa con los usuarios, para ello se han generado tutores virtuales que acompañan al usuario en su preparación, los tutores virtuales deben tener características sociales muy similares a las de los seres humanos (Krämer, 2017), con el fin de que los usuarios se motiven y no pierdan el interés en continuar y finalizar su aprendizaje.

Los tutores virtuales son una poderosa herramienta de la IA que actualmente es utilizada en un sinnúmero de sistemas y aplicaciones educativos, la fortaleza radica en la adquisición automática de nueva información que es clasificada y presentada al usuario para reforzar su aprendizaje. Existen varios mecanismos de recopilación de información que los tutores virtuales deben clasificar antes de ser presentados al usuario para garantizar su aprendizaje, en este trabajo se propone que el tutor virtual esté dotado de un Sistema Clasificador de Aprendizaje.

Sistema Clasificador de Aprendizaje (LCS)

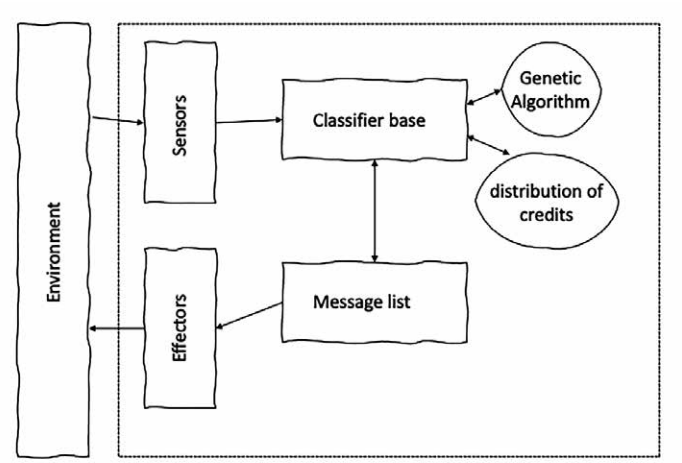
Los sistemas LCS propuestos por John Holland (Holland *et al.,* 2000) son una técnica de aprendizaje por refuerzo supervisado y no supervisado. Este tipo de sistemas recibe entradas codificadas en binario, provenientes del medio ambiente que se alojan dentro de la memoria; estas entradas son llamadas *lista de mensajes.* El LCS debe responder de forma apropiada y realizar una acción que permita al tutor virtual ejecutar una tarea o cumplir con sus objetivos.

John Holland (Holland *et al.,* 2000) describe los LCS como un marco de trabajo (véase figura 3.4) que utiliza un algoritmo genético para analizar el aprendizaje basado en reglas y el par de reglas "condición/acción". El conjunto de reglas que conforman la acción se va formando al utilizar la notación: # símbolo que significa "no importa" o el símbolo ? que corresponde a la notación "se adapta a todos", 0 genera un falso y 1 genera un valor verdadero. Los símbolos # y ? se utilizan indistintamente en la implementación del LCS (véase figura 3.5), ya que ambos símbolos pueden aceptar cualquier valor permitido en el entorno (Holland *et al.,* 2000; Lanzi, Stolzmann y Wilson, 2000).

Tutor inteligente dotado de un LCS

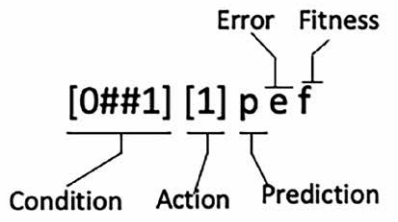
Un LCS es una máquina de aprendizaje que combina aprendizaje supervisado y algoritmos evolutivos para resolver problemas de clasificación o toma de decisiones. En este caso se busca que el tutor virtual este dotado de un LCS para que pueda interactuar con el usuario y trasmitir el conocimiento, mientras mejora su propio conocimiento, tal como lo hacemos los seres humanos.

Figura 3.4. Arquitectura general de un sistema LCS



Fuente: Holland *et al.,* 2000.

Figura 3.5. Reglas condición /acción del LCS



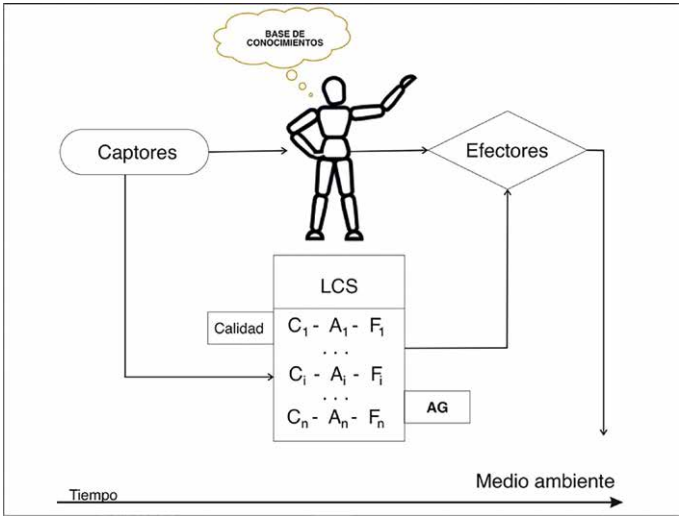
Fuente: Lanzi, Stolzmann y Wilson, 2000.

En la figura 3.6 se presenta la propuesta de la arquitectura de un tutor inteligente dotado de un LCS como mecanismo de seguimiento, evaluación y aprendizaje que permita darle un seguimiento puntual a los usuarios al momento de adquirir un nuevo conocimiento.

La IA de un tutor inteligente dotado de un LCS se basa en la codificación de reglas, que posteriormente son clasificadas de acuerdo con su nivel de importancia, y que mediante un mecanismo evolutivo permite a un tutor inteligente aprender y adaptarse a partir de la interacción con su medio ambiente y con los usuarios.

La arquitectura propuesta permite que los usuarios interactúen con el tutor inteligente, que les proporciona información relevante, responde preguntas y mejora su capacidad de transmitir el conocimiento a medida que recibe retroalimentación del medio ambiente y de los usuarios.

Figura 3.6. Propuesta de la arquitectura de un tutor inteligente dotado de un LCS



Fuente: elaboración propia.

Los componentes que conforman la arquitectura propuesta para el tutor inteligente dotado por el LCS son:

1. Base de conocimientos: el tutor inteligente almacena la información de forma estructurada sobre el tema o conocimientos específicos de un área en particular (por ejemplo: historia, matemáticas, etc.), en donde la base de conocimientos puede responder hechos o eventos importantes sobre el área de interés que se esté abordando, como fechas, lugares, personajes, fórmulas, eventos, etcétera.
2. Sistema Clasificador de Aprendizaje: el LCS emplea un sistema de reglas que clasifica las preguntas del usuario y les asigna respuestas adecuadas. Con el tiempo, el tutor inteligente mejora la clasificación de nuevas preguntas mediante mecanismos evolutivos, comparándolas con el conocimiento ya almacenado. A medida que el tutor inteligente interactúa con los usuarios, este ajusta su base de reglas mediante un algoritmo genético (AG) lo que le permite mejorar su capacidad de respuesta, refinando sus respuestas con base en el historial de interacciones (calidad). Este proceso puede integrar retroalimentación positiva o negativa del usuario para guiar el aprendizaje del tutor inteligente.
3. Interacción a través del tiempo: los usuarios se sumergen en un ambiente virtual (por ejemplo, una reconstrucción 3D de un aprendizaje específico) en el que pueden interactuar directamente con el tutor inteligente, que responde a preguntas específicas sobre el tema que se esté abordando y conforme transcurre el tiempo, las respuestas a las preguntas se vuelven conocimiento adquirido, tanto para el tutor como para el usuario.
4. Ambiente virtual: el entorno virtual simula la experiencia del usuario en un contexto realista, mejorando la comprensión y el interés en el contenido educativo. La inmersión puede incluir elementos visuales y auditivos, lo que hace que la enseñanza sea más atractiva y memorable. La interacción del tutor inteligente en el ambiente virtual le brinda al usuario un acompañamiento durante todo su aprendizaje.

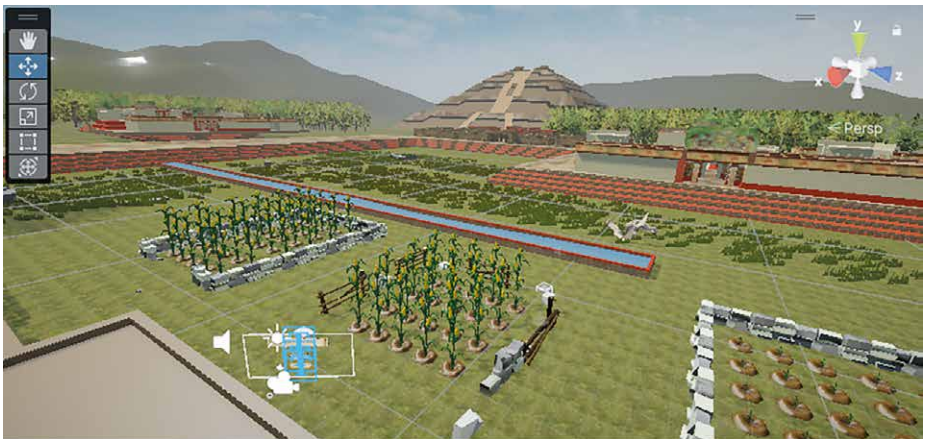
La arquitectura propuesta para el tutor inteligente dotado de un LCS permite que el tutor inteligente adapte el contenido del área de estudio que se esté abordando en función de las preguntas, necesidades y el nivel de conocimiento del usuario, lo que crea una experiencia educativa individualizada. La posibilidad de que el usuario cuente con una interacción directa con el tutor inteligente lo motiva a

explorar más aspectos y a cuestionarse sobre los temas de aprendizaje manteniendo una participación constante. De forma adicional, a medida que el tutor inteligente interactúe con más usuarios, este aprende a clasificar la información eficientemente de acuerdo con las preguntas y respuestas que proporciona a los usuarios.

Implementación de un tutor inteligente en un contexto histórico

La arquitectura propuesta se validó mediante la implementación de un ambiente virtual educativo que aborda los temas de la agricultura prehispánica teotihuacana (https://vianco.uaemex.mx/index.php/vr_teotihuacan/). En la figura 3.7 se ilustra la generación de un ambiente virtual de la zona arqueológica de Teotihuacán, donde se embebió el tutor inteligente con los objetivos de enseñar al usuario sobre la agricultura que se practicaba durante la época prehispánica, así como las técnicas y herramientas que se utilizaban para la producción y cosecha del maíz.

Figura 3.7. Ambiente virtual de la zona arqueológica teotihuacana, Teotihuacán, Estado de México



Fuente: elaboración propia.

Los ambientes virtuales deben proporcionar una sensación vívida del espacio que se representa, así como diferentes maneras de interactuar de forma directa con este. En la figura 3.8, se ilustra la interacción del usuario en un ambiente virtual inmersivo donde el usuario puede ver sus manos mediante dispositivos hápticos, que le permiten interactuar con los objetos existentes en el ambiente virtual, así como mecanismos de comunicación con entidades dentro del ambiente virtual y con el tutor inteligente.

Figura 3.8. Interacción del usuario con el medio ambiente virtual y el tutor inteligente

Fuente: elaboración propia.

El objetivo de este ambiente virtual educativo es enseñar al usuario a utilizar las herramientas que utilizaba la cultura teotihuacana en el cultivo del maíz, esto es, el tutor inteligente mostrará al usuario las técnicas asociadas al cultivo del maíz y su cuidado, posteriormente, el usuario tendrá los conocimientos necesarios para aplicarlos directamente en un espacio destinado dentro del ambiente virtual para que él realice el proceso de cultivo, así como el cuidado para obtener la cosecha de maíz. La figura 3.9 ilustra la interacción dentro del ambiente virtual del usuario con el tutor inteligente, para este caso, se utilizan diferentes medios de comunicación, como el audio y diálogos de texto.

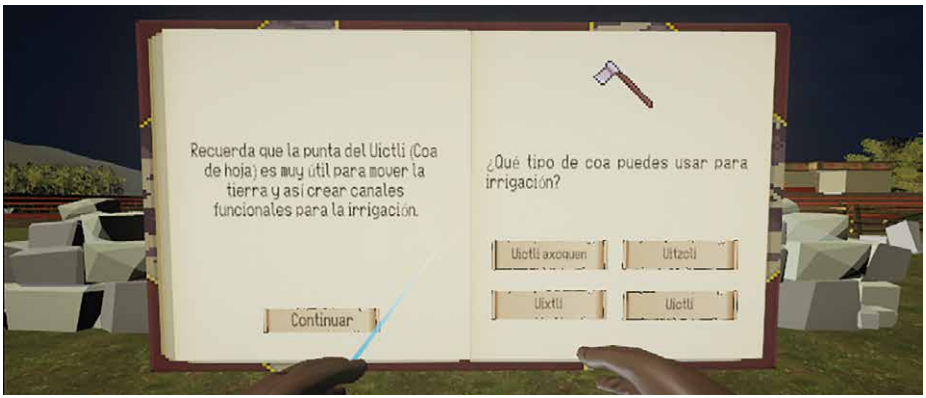
Los tutores virtuales deberán contar con la capacidad de atender las necesidades de cada uno de los usuarios al momento de adquirir nuevos conocimientos e

información, así como diseñar mecanismos de evaluación que validen que el conocimiento fue adquirido y, en caso contrario, replantear la presentación de la información de este conocimiento, de tal forma que el usuario asegure que tiene las habilidades y conocimientos necesarios para aplicarlos directamente a la realización de una tarea o resolución de alguna problemática. La figura 3.10 ilustra uno de los medios de validación que propone el tutor inteligente para verificar que el usuario ha adquirido de forma correcta el conocimiento presentado sobre el cultivo del maíz en la cultura teotihuacana.

Figura 3.9. Tutor inteligente enseñando al usuario el proceso del cultivo de maíz en la época prehispánica

Fuente: elaboración propia.

Figura 3.10. Cuestionario propuesto por el tutor inteligente para validar el conocimiento adquirido por el usuario



Fuente: elaboración propia.

Discusión

Los tutores inteligentes son una herramienta de la IA, que pueden ser utilizados en sistemas educativos, en los que la transferencia de conocimiento hacia el usuario se hace de forma dinámica y personalizada. La posibilidad de que los tutores inteligentes puedan modificar su conocimiento de acuerdo con las características particulares de los usuarios brinda la posibilidad de que el usuario adquiera de forma correcta nuevo conocimiento y que pueda ser aplicado de forma correcta.

Contar con un tutor inteligente capaz de aprender "cómo aprender" posibilita desarrollar sistemas educativos capaces de adaptarse al comportamiento individual de cada usuario, anticipando la manera en que cada usuario es capaz de adquirir el nuevo conocimiento, esto garantiza que los usuarios aprendan según sus habilidades.

La anticipación es un problema abierto que requiere grandes cantidades de información, sobre todo en condiciones de incertidumbre, la novedad de la propuesta radica en dotar al tutor inteligente con un LCS, que permita adaptarse a las necesidades particulares de cada usuario.

Problemas emergentes en la clasificación de información y el manejo de grandes cantidades de datos son abordados y tratados de forma directa en técnicas

como ciencias de datos y percepción o visión computacional. En el siguiente capítulo trataremos la problemática de clasificación de información para el reconocimiento facial.

Referencias

Esteve Zarazaga, J. M. *et al.* (2001). El profesorado de secundaria: hacia un nuevo perfil profesional para enfrentar los problemas de la educación contemporánea. *Revista Fuentes.* https://revistascientificas.us.es/index.php/fuentes/article/view/2730

Holland, J. H., Booker, L. B., Colombetti, M., Dorigo, M., Goldberg, D. E., Forrest, S., Riolo, R. L., Smith, R. E., Lanzi, P. L., Stolzmann, W. *et al.* (2000). What is a learning classifier system. *Learning Classifier Systems,* pp. 3-32. DOI: https://doi.org/10.1007/3-540-45027-0_1

Krämer, N. C. (2017). The immersive power of social interaction: Using new media and technology to foster learning by means of social immersion. *Virtual, augmented, and mixed realities in education,* pp. 55-70. DOI: 10.1007/978-981-10-5490-7_4

Lanzi, P. L., Stolzmann, W. y Wilson, S. W. (2000). Learning classifier systems: from foundations to applications, *Science and Business Media,* núm. 1813.

Nedungadi, P. y Remya, M. S. (2015). Incorporating forgetting in the personalized, clustered, bayesian knowledge tracing (pc-bkt) model. *International Conference on Cognitive Computing and Information Processing (CCIP),* pp. 1-5. DOI: 10.1109/CCIP.2015.7100688

Richards, J. (2017). Infrastructures for immersive media in the classroom. *Virtual, augmented, and mixed realities in education*, Springer Nature, pp. 89-104.

Shute, V., Rahimi, S. y Emihovich, B. (2017). Assessment for learning immersive environments. *Virtual, augmented, and mixed realities in education*. Springer Nature, pp. 71-87.

DETECCIÓN DEL ROSTRO HUMANO EN IMÁGENES 3D

Marcelo Romero Huertas
Facultad de Ingeniería, UAEMEX

Juan Paduano Salinas
Facultad de Ingeniería, UAEMEX

Resumen

La detección del rostro humano es un tema estudiado por los investigadores en visión por computadora durante décadas. A pesar de ello, las técnicas actuales de detección de rostros en imágenes de profundidad aún dependen de la posición de la persona y se han obtenido resultados satisfactorios en imágenes que solo contienen el rostro humano. En este sentido, el desarrollo logrado en la IA abre una alternativa para la detección de rostros humanos en imágenes 3D, utilizando técnicas avanzadas de procesamiento de imágenes y aprendizaje automático para identificar y reconocer características faciales en entornos tridimensionales.

En este capítulo se investigan de forma experimental tres técnicas clave de la detección de rostros. Posteriormente, se presenta la novedad de esta investigación, que consiste en detectar y localizar cada rostro presente en una imagen 3D con independencia del número de sujetos y su distancia de la cámara. Para ello, se presenta un enfoque sencillo y eficaz que consta de cuatro pasos. Se integra una base de datos de imágenes 3D utilizando la cámara Kinect One™, variando posiciones de uno a ocho sujetos diferentes en la escena. Además de la base de datos de imágenes 3D, este enfoque de detección de rostros se ha evaluado con bases de datos de última generación, Face Recognition Grand Challenge (FRGC) y CurtinFaces, que contienen un sujeto por imagen. Los resultados que se presentan son motivadores, considerando un error de localización de 0 a 16 mm se obtienen resultados experimentales del 100, 98 y 91% para FRGC, CurtinFaces y nuestra base de datos, respectivamente.

Introducción

La detección de rostros es el primer paso en casi todas las aplicaciones de procesamiento de rostros, en que el rostro se localiza y se extrae de una imagen entrante antes de un análisis específico (Hjelmas y Low, 2001; Bowyer, Chang y Flynn, 2006; Zafeiriou, Zhang y Zhang, 2015). Sin embargo, la detección de rostros a partir de una sola imagen es una tarea desafiante debido a la variabilidad en escala, ubicación, orientación (vertical, rotada) y pose (frontal, perfil). Asimismo, las expresiones faciales, la oclusión y las condiciones de iluminación cambian la apariencia general del rostro (Bowyer, Chang y Flynn, 2006; Yang, Kriegman y Ahuja, 2002; Yang y Ahuja, 2001; Pears, Liu y Bunting, 2012; Paduano, Romero y Muñoz, 2015). Es evidente que la detección de rostros es el primer paso en cualquier sistema automatizado que resuelva problemas sobre áreas de reconocimiento facial. Sin embargo, la mayoría de los métodos de procesamiento de rostros existentes asumen que solo un rostro humano está presente en una imagen o en un video (Bowyer, Chang y Flynn, 2006; Zafeiriou, Zhang y Zhang, 2015; Yang y Ahuja, 2001; Ariz *et al.,* 2016). Aunque los algoritmos de detección 2D han alcanzado un nivel aceptable en la detección de rostros tienen muchas dificultades porque el mundo 3D se proyecta sobre una imagen 2D, lo que genera ambigüedades y pérdida de información de profundidad (Pears, Liu y Bunting, 2012). Por otro lado, las imágenes 3D proporcionan información explícita sobre la forma y se consideran resistentes a la iluminación y las variaciones de pose (Bowyer, Chang y Flynn, 2006; Pan, Zhu y Xia, 2013) utilizando descriptores de características heterogéneos y selección de características en combinación con Adaboost (Yin *et al.,* 2015; Hu, Hu y Maybank, 2018), aunque este es un método de detección de rostros robusto y eficiente, depende de la posición frontal. Otras investigaciones de detección de rostros dependientes de la pose que utilizan imágenes 2D son: Subburaman y Marcel, 2013; Aghaei, Dimiccoli, Radeva, 2016; Bhandarkar y Luo, 2009; Wang y Ji, 2007). La detección de rostros es necesaria para la aplicación de procesamiento de rostros, por ejemplo, reconocimiento de rostros (Wagner *et al.,* 2012; Kumar, Datta, Kumar, 2015; Kakadiaris *et al.*, 2016), seguimiento de rostros (Aghaei, Dimiccoli y Radeva, 2016; Bhandarkar y Luo, 2009; Smeulders, 2019; Smeulders *et al.,* 2014),), estimación de pose (Wohlhart y Lepetit, 2015), reconocimiento de expresiones y gestos faciales (Valstar *et al.,* 2012). Desafortunadamente, diferentes factores como

la posición del sujeto, el sensor 3D, los componentes estructurales, la oclusión, la resolución de la superficie y la expresión facial hacen que la detección de rostros sea un problema complicado (Bowyer, Chang y Flynn, 2006; Pears, Liu y Bunting, 2012; Kumar, Datta, Kumar, 2015; Yang *et al.,* 2014; Chellappa, Wilson y Sirohey, 1995).

Aunque la detección de rostros es esencial para muchas aplicaciones de procesamiento de rostros ha recibido poca atención, en especial, cuando se utilizan datos 3D. En este capítulo, se propone un enfoque innovador para la detección de rostros en 3D, utilizando una vista de escenas con variaciones en el número de personas y la distancia al sensor.

Trabajo relacionado

Hay varios artículos sobre aplicaciones de procesamiento de rostros, sin embargo, solo unos pocos están dedicados específicamente a la detección de rostros. Esta sección analiza trabajos relevantes relacionados con la detección de rostros en imágenes 3D.

Mian, Bennamoun y Owens (2006) propusieron un método de detección de la punta de la nariz mediante el uso de contornos cortados horizontalmente a lo largo de una imagen facial. Cada contorno cortado está formado por *n* vértices. Para cada vértice en cada contorno cortado, se estimaron círculos de radio constante con centro en cada vértice. A continuación, se formaron triángulos circunscritos en todos los círculos utilizando tres vértices, un vértice es el centro del círculo y dos vértices de intersección ubicados hacen coincidir el contorno del círculo y los puntos en el corte. Después de la comparación de la altitud del triángulo, solo se selecciona un triángulo en cada corte como candidato para la punta de la nariz, el que tiene la altitud máxima desde el centro del círculo. Se utilizó el Consenso de Muestra Aleatoria (RANSAC) (Trucco y Verri, 1998) junto a una mayor eliminación de valores atípicos y de todos los candidatos de punta de nariz restantes, se toma como punta de nariz la que tiene la altitud máxima. Finalmente, se extrae el rostro utilizando un radio preestablecido. De esta forma se informó de una precisión del 98.3% en la detección de rostros utilizando la base de datos FRGC.

Colombo, Cusano y Schettini (2006) propusieron una técnica basada en el análisis de curvatura para la detección de rostros. Comenzaron el proceso utilizando

una imagen de rango que contiene una sola cara, que es una representación de las coordenadas (*x, y, z*) de la imagen 3D en una ubicación (*i, j*). La ubicación de cada punto 3D se expresó con respecto al sistema de referencia del sensor 3D. A continuación, a partir de las imágenes de rango, se calcularon la curvatura media y gaussiana mediante la primera y segunda derivadas en *x* y *y*. Consideraron una aproximación polinómica bi-cuadrática de la superficie y los coeficientes se obtuvieron usando mínimos cuadrados ajustando los puntos en una vecindad de *x* y *y*. El objetivo para calcular las curvaturas media y gaussiana es utilizar la clasificación HK para obtener cuatro tipos de segmentos: convexo, cóncavo y dos regiones de silla. El resultado puede contener muchos candidatos para rasgos faciales. Si no se detecta ninguna nariz o menos de dos ojos candidatos, se supone que no hay ningún rostro presente en la escena. Finalmente, si se detecta una cara, el resultado es una lista que contiene su ubicación y área facial. Se experimentó con estos datos recopilados utilizando una cámara Minolta™ y se logró un 96.85% de éxito en la detección de rostros.

Segundo *et al.* (2007) presentaron un método para extraer la región del rostro de una sola persona. También utilizaron imágenes de rango generadas a partir de una nube de puntos. Su algoritmo de segmentación de rostros consta de dos etapas principales: primero, se utiliza la localización de regiones homogéneas en la imagen de entrada mediante agrupamiento y detección de bordes. En segundo lugar, se identifican los candidatos cuando pertenecen a la región de la cara, luego se delimita esta región mediante una elipse utilizando la transformada de Hough. Aplicaron el algoritmo K-Means con $k = 3$ para la segmentación de imágenes. Entonces, la imagen de entrada se segmentó en tres regiones principales identificadas como fondo, cuerpo y rostro. Sin embargo, este paso por sí solo no es suficiente para la segmentación de la cara, por lo que es necesario extraer la región de la cara mediante la detección de bordes. Después de realizar la detección de regiones y bordes, que se puede hacer en paralelo, combinan las dos imágenes resultantes y, finalmente, mediante una operación lógica *and,* se obtuvo la región facial. Informaron de una detección de rostros del 99.95% utilizando la base de datos FRGC.

Nair y Cavallaro (2009) presentan un marco preciso y sólido para la detección y segmentación de rostros utilizando la localización de puntos de referencia y realizando un registro fino de mallas faciales para ajustar un modelo facial (polígonos

y vértices). Este modelo se basa en un modelo de distribución de puntos (PDM) 3D que se ajusta sin depender de información de textura, pose u orientación. El ajuste comienza utilizando ubicaciones candidatas en la malla, que se extraen de mapas de características basados en curvatura de bajo nivel. La detección de rostros se hace clasificando las transformaciones entre los puntos del modelo y los vértices candidatos en función del límite superior de la desviación de los parámetros del modelo medio. El rendimiento de la detección de rostros se evalúa en una base de datos de imágenes de rostros y sin rostros logrando una precisión del 99.6%. También demostraron detección y segmentación de rostros con diferentes escalas y poses.

A partir de la literatura revisada con anterioridad se han investigado experimentalmente tres técnicas clave de detección de rostros (Paduano, Romero y Muñoz, 2015) utilizando bases de datos de última generación: Face Recognition Grand Challenge y CurtinFaces para una comprensión profunda. Los hallazgos y el conocimiento de esta investigación experimental permiten proponer un nuevo enfoque de detección de rostros (Paduano, Romero y Valdovinos, 2016). Además, se pueden identificar una serie de investigaciones relacionadas con el procesamiento facial utilizando imágenes 2D y 3D. Por ejemplo, Peng y Bennamoun (2011) proponen un método de detección de la punta de la nariz que tiene tres características principales. En primer lugar, no requiere formación y no depende de ningún modelo en particular. En segundo lugar, puede abordar posturas tanto frontales como no frontales. Finalmente, es bastante rápido y solo requiere unos segundos para procesar una imagen de 100 a 200 píxeles (tanto en dimensiones *x* como *y*). Chang, Bowyer y Flynn (2006) establecieron dos umbrales para los dos mapas de curvatura para buscar la punta de la nariz, los dos ojos y la cresta de la nariz, utilizaron una base de datos de 4 000 imágenes de 449 sujetos diferentes. Hutton, Buxton y Hammond (2003) utilizan un ajuste híbrido de punto más cercano iterativo (ICP) y modelo de forma activa (ASM) para el registro no rígido de un modelo de superficie densa en caras 3D. Este método no requiere textura, por lo que impone restricciones en la orientación de la cara y no es invariante de escala.

Por otro lado, Nanni *et al.* (2014) presenta un detector de rostros basado en el algoritmo de Viola y Jones. Soldera, Behaine y Scharkanski (2015) proponen un método de reconocimiento facial en imágenes 2D basado en proyecciones de imágenes faciales de alta dimensionalidad en dimensiones bajas. En su estudio

de reconocimiento facial se explica que sigue siendo una tarea difícil, ya que las imágenes 2D del rostro pueden verse afectadas por cambios en la escena, por ejemplo: posición del rostro, expresiones faciales o iluminación. Su método de reconocimiento es una modificación de la técnica introducida por Cai *et al.* (2006), que propone un método de reconocimiento facial basado en la apariencia, denominado caras laplacianas ortogonales (OLPP). Zhang, Zhang y Ha (2008) proponen un método para detectar rostros en imágenes 2D basado en el análisis de componentes principales (PCA) y la máquina de vectores de soporte (SVM). Primero, filtran áreas que potencialmente son una cara utilizando una función de distribución estadística del histograma local. Luego, las caras candidatas se reducen utilizando un clasificador SVM. Chen *et al.* (2009) proponen un método de detección de rostros utilizando un modelo de rostro de media plantilla 2D. En esta investigación explican que la detección de rostros es un proceso importante para el reconocimiento facial. Su modelo es una plantilla de rostro promedio que contiene ojos, nariz, boca y parte de la mejilla. La detección de rostros se hizo utilizando un método de coincidencia de plantillas para determinar la posición del rostro en la imagen. Su análisis teórico muestra que la plantilla de rostro promedio puede reducir las posibilidades de densidad local y puede adaptarse a imágenes de rostros laterales en ángulo, lo que mejora la precisión de la detección de rostros en posición lateral.

Belahcene, Chouchane y Mokhtari (2014) proponen un sistema de reconocimiento facial (FRS) de cinco etapas: preprocesamiento y detección, rama, fusión de características, extracción y clasificación de características. La primera etapa, consiste en detectar el rostro mediante proyección integrada de curvas horizontales y verticales. En la segunda etapa se divide el rostro mediante parches faciales que consisten en la nariz y los ojos. En la tercera etapa se combina la información de la imagen de la profundidad de la imagen 2D, que es utilizada por PCA y el modelo mejorado Fisher (EFM) para extraer características. Estos se utilizaron para la etapa de clasificación utilizando dos métodos: medidas basadas en distancia y una máquina de vectores de soporte (SVM). Maes *et al.* (2010) presentan un algoritmo SIFT adaptado para superficies 3D (llamado meshSIFT) y sus aplicaciones a la normalización y el reconocimiento de poses faciales 3D. Su algoritmo permite una detección confiable del espacio de escala como ubicaciones de características locales. Luego, el algoritmo

meshsift describe la vecindad de cada espacio de escala en un vector de características que consta de histogramas concatenados de índices de forma y ángulos. Después, los vectores de características se hacen coincidir de manera confiable comparando el ángulo en el espacio de características.

Detección del rostro en imágenes 3D con una persona

Esta sección presenta un análisis experimental de tres enfoques clave de detección de rostros utilizando imágenes 3D con una persona: análisis de curvatura propuesto por Colombo, Cusano y Schettino (2006), el análisis por cortes propuesto por Mian, Bennamoun y Owens (2006), y el análisis por segmentación propuesto por Segundo *et al.* (2007).

Análisis de curvatura

Colombo, Cusano y Schettino (2006) utilizan el análisis de curvatura para detectar el rostro humano contenido en una imagen 3D, ejecutando los siguientes pasos:

Paso 1. Las imágenes de profundidad se calcularon a partir de su respectiva nube de puntos 3D utilizando las ecuaciones 4.1 y 4.2.

$$f \rightarrow U; definido\ en\ un\ conjunto\ abierto\ U\ \subseteq R^2 \tag{4.1}$$

$$S = (x, y, z) | (x, y)\ \in U; z\ \in R; f(x, y) = z \tag{4.2}$$

Paso 2. Se aplicó un filtro Gaussiano (ecuación 4.3) a la imagen de profundidad de entrada para descartar fluctuaciones de alta frecuencia en la superficie, mientras que los rasgos faciales destacados, como los ojos y la nariz, aún se distinguen claramente. Cuando el filtro se aplica a toda la imagen, algunas regiones como esquinas o contornos de caras no se descartan debido a la distorsión radial.

$$\frac{1}{16} * \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix} \tag{4.3}$$

Paso 3. La curvatura media se observa en la ecuación 4.4 y gaussiana, en la 4.5, se calcularon mediante la primera (ecuación 4.6) y segunda (ecuación 4.7) derivadas de cada punto del mapa de profundidad como en Trucco y Verri (1998).

$$H(x,y) = \frac{\left(1 + f_y^2\right) f_{xx} - 2 f_x f_y f_{xy} + (1 + f_x^2) f_{yy}}{2(1 + f_x^2 + f_y^2)^{\frac{3}{2}}} \tag{4.4}$$

$$K(x,y) = \frac{f_{xx} f_{yy} - f_{xy}^2}{(1 + f_x^2 + f_y^2)^2} \tag{4.5}$$

Donde *fx, fy, fxy, fxx, fyy* son las derivadas primera y segunda de *f* en (x, y). Entonces, un rostro se representa inicialmente como una imagen de rango. Como solo tenemos una representación discreta de *S,* también se calcularon las derivadas parciales para cada punto del mapa de profundidad:

$$f'(x_0) = \frac{f(x_0 + h) - f(x_0 - h)}{2h} + O(h^2) \tag{4.6}$$

$$f''(x_0) = \frac{f(x_0+h) - 2f(x_0) + 2f(x_0-h)}{h^2} + O(h^2) \tag{4.7}$$

considerando que x_0 es un punto (x, y) en la imagen del rango, h es un número entero mayor que 0 y $O(h^2)$ es el error de truncamiento, causado al detener el polinomio aproximación de segundo orden, que tiende a cero.

Cuadro 4.1. Clasificación *HK*

	$K<0$	$K=0$	$K>0$
$H<0$	Cóncavo hiperbólico	Cóncavo cilíndrico	Cóncavo elíptico
$H=0$	Simétrico hiperbólico	Plano	Imposible
$H>0$	Convexo hiperbólico	Convexo cilíndrico	Convexo elíptico

Fuente: Besl y Jain, 1986.

Paso 4. Se hace la clasificación *HK* con base en los signos de curvatura media y gaussiana, haciendo esto, se pueden identificar puntos elípticos convexos cuando $K > 0$ y $H > 0$ y esos puntos podrían representar una nariz (véase cuadro 4.1).
Paso 5. Se hace un proceso de umbralización con base en los valores de curvatura (como se observa en la ecuación 4.8). Las regiones con mayor curvatura se aíslan (como se muestra en la figura 3), como candidatos a nariz, mientras que las áreas con valores de curvatura bajos se descartan.

$$|H(u,v)| \geq T_h \qquad |K(u,v)| \geq T_k; \tag{4.8}$$

donde T_h y T_k son los umbrales estimados:

$$T_k = \frac{k_{máx} - k_{mín}}{4} \tag{4.9}$$

$$T_h = \frac{h_{máx} - h_{mín}}{4} \tag{4.10}$$

Paso 6. Finalmente, se reduce el número de candidatos filtrando cada candidato elíptico convexo, considerando la curvatura media (ver ecuación 4.11) y gaussiana (ecuación 4.2)

$$\overline{H_l} \geq \overline{H_{mín}} \tag{4.11}$$

$$\overline{K_l} \geq \overline{K_{mín}} \tag{4.12}$$

Análisis por cortes

Para la detección del rostro humano contenido en una imagen 3D, Mian, Bennamoun y Owens (2006) hacen un análisis por cortes, que consiste en los siguientes pasos:

Paso 1. La imagen 3D se secciona horizontalmente en varios cortes utilizando un umbral *dv*. Donde para cada corte, los picos se eliminan usando un umbral *dt*, que se calcula automáticamente mediante la expresión 4.13:

$$dt = \mu + 0.6\sigma \tag{4.13}$$

donde μ es la distancia media entre los ocho puntos vecinos con desviación estándar σ.

Paso 2. Después de quitar los picos, la imagen 3D puede resultar en agujeros. Esos agujeros se rellenaron mediante interpolación cúbica. Luego, se crearon círculos centrados en múltiples intervalos horizontales *dh* en cada corte para una segmentación de región. En cada área segmentada se inscribió un triángulo, desde el centro del círculo, y dos puntos de intersección que coinciden con el círculo y los puntos segmentados del corte. Los círculos se crearon con un radio de 2 cm asociado al tamaño medio de la nariz desde la punta hasta el final de la nariz.

Paso 3. En cada corte debe haber un triángulo de máxima altitud. Este triángulo puede considerarse como una posible punta de nariz. Se le asigna un valor de confianza igual a la altitud del triángulo, para hacer una mayor comparación con los otros triángulos de altitud máxima de otros sectores.

Paso 4. Después de obtener los valores de confianza del paso 3, asociados a todos los cortes, se identifica un punto candidato a punta de nariz por corte. Esos candidatos

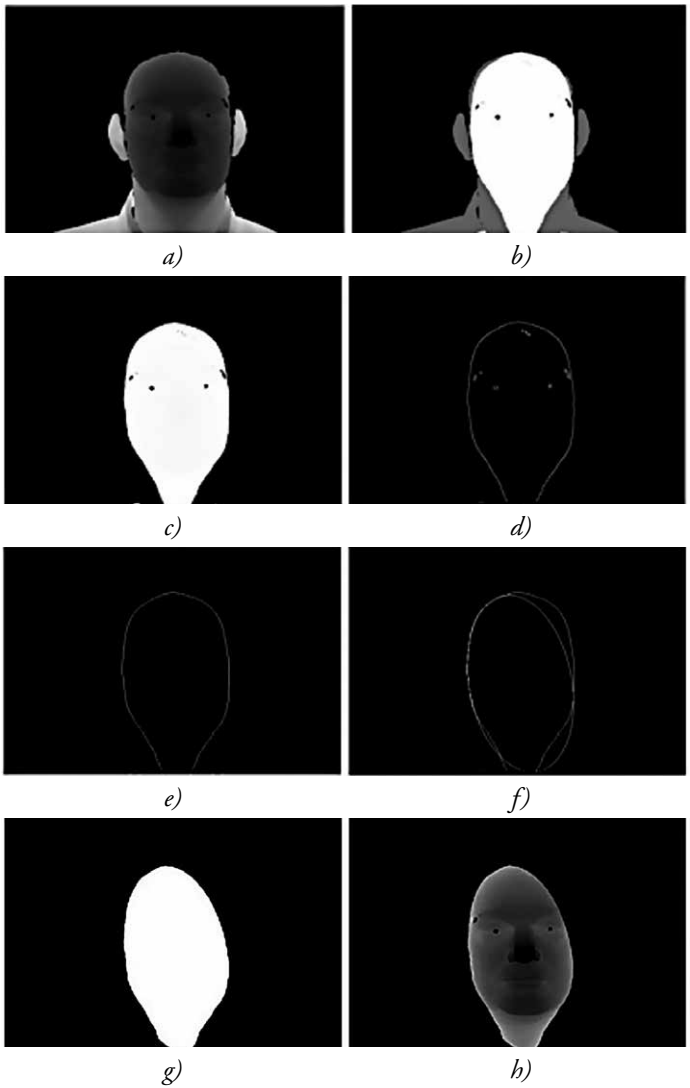
fueron analizados para identificar la cresta de la nariz. Los puntos candidatos que no corresponden a la cresta de la nariz se consideraron valores atípicos y se eliminaron utilizando RANSAC (Trucco y Verri, 1998).
Paso 5. Finalmente, se evaluaron los pocos puntos candidatos y se definió el de máxima altitud como punta de la nariz. Luego, se extrae toda la cara desde la punta de la nariz detectada.

Análisis por segmentación

Segundo *et al.* (2007) hacen un análisis por segmentación para la detección del rostro humano contenido en una imagen 3D, que consiste en los siguientes pasos:

Paso 1. Las imágenes de profundidad de entrada (véase figura 4.1-a) para este algoritmo se generaron a partir de una nube de puntos 3D utilizando las ecuaciones (4.1) y (4.2). El proceso completo consta de siete etapas: segmentación K-Means, extracción de región facial, operación de Sobel, cierre, detección de elipse, binarización y segmentación (véase figura 4.1).
Paso 2. El algoritmo K-Means (Krishna y Murti, 1999) se aplica estableciendo $k = 3$, por lo que la imagen de entrada se segmentó en tres regiones principales: fondo, cuerpo y rostro (véase figura 4.1-b). Todo el fondo está en negro, el cuerpo en gris y la región de interés, la cara, en blanco (véase figura 4.1-c), solo la región blanca se extrae y se utiliza en los siguientes pasos.
Paso 3. La detección de bordes se aplica utilizando el operador Sobel y la región de la cara ahora está representada por bordes (véase figura 4.1-d).
Paso 4. Luego, solo se mantiene el borde mayor (véase figura 4.1-e) al que se le ajusta una elipse mediante transformada de Hough. La imagen resultante es una imagen binaria limpia (véase figura 4.1-f) con la posible forma de la cara.

Figura 4.1. Ilustración de nuestra implementación: *a)* imagen de rango, *b)* resultado del algoritmo K-means, *c)* extracción de región facial, *d)* resultado del operador Sobel, *e)* resultado del proceso de cierre, *f)* detección de elipse con transformación de Hough, *g)* imagen binaria después de la normalización y *h)* segmentación final



Fuente: basada en Segundo *et al.*, 2007.

Paso 5. Finalmente, para obtener un mapa de bordes con la cara detectada y segmentada, en el último paso se aplicó a la imagen resultante un proceso de cierre. Al utilizar una operación lógica *and* entre esta imagen y la imagen de profundidad de entrada, se obtiene la región de la cara.

Resultados experimentales

Para comparar las tres técnicas descritas en las secciones anteriores, se utilizan las bases de datos CurtinFaces (Li *et al.,* 2013) y FRGC (Phillips, 2005). CurtinFaces es una base de datos que se integró utilizando la cámara Kinect 360, por lo que resultan imágenes 3D similares a las recopiladas para propósitos de esta investigación. Por otra parte, se experimenta con la base de datos del FRGC, debido a que es un referéndum en las investigaciones de procesamiento facial.

Cuadro 4.2. Porcentaje de detección correcta del rostro humano en imágenes 3D utilizando las bases de datos del FRGC y CurtinFaces

Técnica	FRGC	CurtinFaces
Análisis de curvatura	98.48 %	75.18 %
Análisis por cortes	92.74 %	57.47 %
Análisis por segmentación	97.19 %	46.88 %

Se implementó cada enfoque de detección de rostros mediante el desarrollo de funciones de propósito específico o utilizando bibliotecas estándar disponibles en Matlab versión 15. Para los tres algoritmos de detección de rostros, el rendimiento se calcula contando aquellas áreas faciales que contienen la punta de la nariz según la imagen de referencia verdadera (Creusot, Pears y Austin, 2010). Para la técnica de corte utilizamos solo la punta de la nariz para detectar una cara. Para las técnicas de curvatura y segmentación utilizamos las comisuras de los ojos y la punta de la nariz. El cuadro 4.2 resume el rendimiento obtenido.

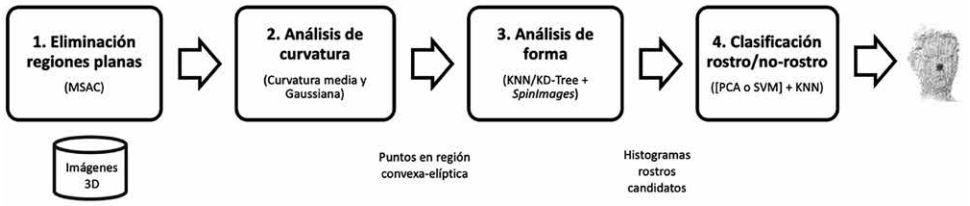
Detección del rostro en imágenes 3D con más de una persona

En esta sección se reporta la investigación experimental para detectar rostros humanos en imágenes de una profundidad que contienen más de una persona en la escena.

Técnica de detección

Si se tienen en cuenta los resultados y el conocimiento obtenidos del análisis experimental descrito en la sección "Detección del rostro en imágenes 3D con una persona", proponemos un enfoque novedoso que es capaz de detectar cada rostro presente dentro de una imagen que contiene a más de una persona en la escena. Como se ilustra en la figura 4.2, este enfoque consta de cuatro pasos: *a)* eliminación de regiones planas, *b)* análisis de curvatura, *c)* análisis de forma (selección de candidatos para la punta de la nariz), y *d)* clasificación rostro /no-rostro. Una descripción detallada de estos pasos es la siguiente.

Figura 4.2. Procedimiento experimental para detectar cada rostro en una imagen 3D que consta de cuatro pasos principales



Fuente: elaboración propia.

Paso 1. Para cada imagen 3D, las áreas planas se eliminaron utilizando el estimador por consenso de muestra (MSAC). MSAC es una modificación de la técnica RANSAC, donde el objetivo es ajustar planos en la nube de puntos usando una distancia *d;* al establecer el plano, se calcula la distancia de los puntos al plano para obtener dos conjuntos de puntos, valores internos y valores atípicos.

Los valores interiores son los vértices que se ajustan al plano calculado y los valores atípicos son aquellos cuya distancia al plano es mayor que *d*. Nuestra implementación de la técnica MSAC consta de cinco pasos:

1. La entrada del algoritmo MSAC es una nube de puntos y una distancia *d*.
2. Un plano $P_1 X + P_2 Y + P_3 Z + P_4 = 0$ se ajusta a una parte de la imagen 3D (x_i, y_i, z_i) con la transformación (4.14):

$$\begin{bmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & z_n & 1 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \end{bmatrix} = A\overline{p} \tag{4.14}$$

3. Se calculan los parámetros óptimos. Definimos los parámetros óptimos como los valores mínimos de ajuste del plano en la ecuación (4.15):

$$\overline{p}opt = argmin||A\overline{p}||^2 \tag{4.15}$$

4. Se calcula el error de estimación de la distancia de cada punto *(x, y, z)* al plano $P_1 X + P_2 Y + P_3 Z + P_4 = 0$:

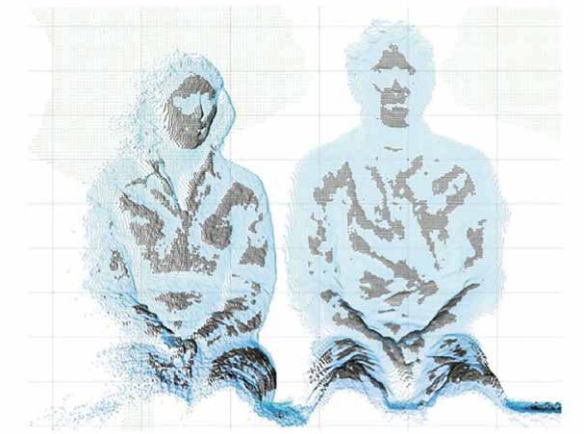
$$e^2 = \frac{([x\ y\ z\ 1]\overline{p}opt)^2}{P_1^2 + P_2^2 + P_3^2} \tag{4.16}$$

5. Seleccione los *inliers* usando una distancia *d*. Si la distancia de un punto es menor o igual que el valor de *d* se considera valor interno, en caso contrario se considera valor atípico. Seleccionamos los valores atípicos porque son puntos que no son regiones planas.

Paso 2. Luego se utiliza el análisis de curvatura para detectar puntos elípticos convexos. Los pasos para implementar el análisis de curvatura son:

1. Se comienza con una imagen de rango que es una representación de las coordenadas (*x, y, z*) de la imagen 3D en una ubicación (*i, j*). Las imágenes de profundidad se calculan a partir de las respectivas nubes de puntos 3D utilizando las ecuaciones (4.1) y (4.2).
2. Para cada punto dentro del mapa de profundidad, se calcula la curvatura media (ecuación 4.4) y gaussiana (ecuación 4.5).
3. Luego, analizando los signos de la media y la curvatura gaussiana, se hace lo que se llama clasificación HK (Besl y Jain, 1986). Para este caso, solo se han seleccionado los puntos elípticos convexos, como se ilustra en la figura 4.3. En este punto, se tienen principalmente esos puntos alrededor del cuerpo y la cabeza del sujeto.

Figura 4.3. Los vértices negros son puntos elípticos convexos



Fuente: elaboración propia.

Paso 3. Para cada vértice de la nube de puntos 3D, se calcularon imágenes de giro (*SpinImages*), según lo prescrito por Johnson y Hebert (1998), mediante los siguientes pasos:

1. Seleccione un punto *p* en la nube de puntos.
2. Calcule normal de *p* usando un radio *r* para seleccionar vecinos de *p*.

3. Calcule los valores α y β, utilizando las ecuaciones (4.17) y (4.18), respectivamente,

$$\propto = \sqrt{||x - p||^2 - (\bar{n} \cdot (x - p))^2} \tag{4.17}$$

$$\beta = (\bar{n} \cdot (x - p)) \tag{4.18}$$

donde *x* es el punto (*x, y, z*) de giro y $\bar{n}$ es el vector normal al plano local en *x*. Cada coordenada α y β contribuye a su respectivo histograma de imagen de giro (*i, j*):

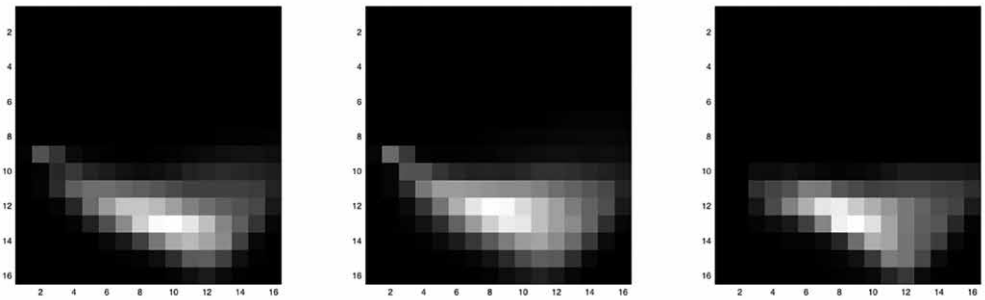
$$\mathrm{i} = [\frac{\beta \text{ máx} - \beta}{\text{bin}}], \mathrm{j} = [\frac{\propto}{\text{bin}}] \tag{4.19}$$

donde *bin* es una constante predefinida, que en esta investigación se calculó experimentalmente como 0.2.

Para acelerar el proceso de selección para crear un histograma de imagen de giro, se utilizaron los algoritmos vecinos más cercanos y árbol KD. Para calcular los vecinos más cercanos de cada vértice se utilizó el árbol KD para dividir el espacio que organiza los puntos en un espacio euclidiano de *k* dimensión. Con esta partición, se obtienen los vecinos de cada vértice usando K-NN, se usó una distancia de 8 cm para la selección de vecinos para generar el histograma de la imagen de giro. La figura 4 muestra histogramas de imágenes de giro calculados a partir de tres bases de datos diferentes en el nivel de la punta de la nariz.

Paso 4. La clasificación rostro /no-rostro se realizó utilizando clasificadores PCA / SVM y K-NN. Para la clasificación PCA comparamos los histogramas de imágenes de giro de los candidatos a punta de nariz restantes. Para esto, se observa un histograma [*pxq*] como un vector de características [*mx1*] en un espacio dimensional *m = pq*. Se definieron conjuntos separados de entrenamiento y prueba y se utilizó un esquema de vecino más cercano propuesto por Pears, Liu y Bunting (2012) basados en la distancia de Mahalanobis de la siguiente manera:

Figura 4.4. Ejemplos de histograma de imagen de giro:
***a)* FRGC, *b)* rostros Curtin, y *c)* BDUAEMEX**



a) *b)* *c)*

Fuente: elaboración propia.

Entrenamiento PCA

1. Para el conjunto de *n* imágenes de entrenamiento, *xi, i = 1…n,* donde cada cara de entrenamiento se representa como un vector de columna *m*-dimensional, como se observa en la ecuación (4.20), apile los *n* vectores de caras de entrenamiento para construir la matriz de datos de entrenamiento [*nxm*]: la ecuación (4.21), donde cada vector de columna representa el histograma de la imagen de giro. En este experimento se utilizaron histogramas [16 × 16] de 200 puntas de la nariz.

$$x = [x1, \ldots, x_m]^T \tag{4.20}$$

$$\mathrm{X} = \begin{bmatrix} X_1^T \\ \vdots \\ X_n^T \end{bmatrix} \tag{4.21}$$

2. La media del conjunto de entrenamiento se calcula como se muestra en la ecuación (4.22), para formar la matriz de datos de entrenamiento de media cero se utiliza la ecuación (4.23) donde *J(n, 1)* es una matriz de unos $n \times 1$.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{4.22}$$

$$X_{zmean} = X - J(n,1)\bar{x}^T \tag{4.23}$$

3. Haga una descomposición propia estándar. En la literatura existen varias variantes de clasificación o reconocimiento facial 3D basado en PCA (Pears, Liu y Bunting, 2012; Colombo, Cusano y Schettini, 2006; Bouzalmat, Kharroubi y Zarghili, 2014; Gumus *et al.,* 2010). Usamos la descomposición de valores singulares (SVD) directamente en la matriz de datos de entrenamiento de media cero *n x m, Xzmean.* La ventaja de utilizar SVD es que a menudo puede proporcionar una estabilidad numérica superior, en comparación con los algoritmos de descomposición propia; además, el almacenamiento requerido para una matriz de datos suele ser mucho menor que el de una matriz de covarianza (Pears, Liu y Bunting, 2012). La SVD es la ecuación (4.24) donde U y V son matrices ortogonales de dimensión *nxn* y *mxm,* respectivamente, y S es una matriz *n x m* de valores singulares a lo largo de su diagonal.

$$USV^T = X_{zmean} \tag{4.24}$$

4. Seleccione el número de dimensiones del subespacio para la proyección. Este es el paso de reducción de dimensionalidad y, generalmente, se hace analizando la relación entre la varianza acumulada asociada con las primeras *k* dimensiones del espacio de la imagen rotada y la varianza total asociada con el conjunto completo de *m* dimensiones en ese espacio.

5. Proyecte los datos de entrenamiento establecidos en el subespacio *k*-dimensional, se utiliza la ecuación (4.25) donde *Vk* es una matriz *mxk* que contiene los primeros *k* vectores propios de *V* y $\tilde{X}$ es una matriz *nxk* de *n* caras de entrenamiento en el subespacio *k*-dimensional.

$$\tilde{X} = X_{zmean}\, V_k \tag{4.25}$$

Prueba PCA

Una vez que se completa la fase de entrenamiento de PCA anterior es sencillo implementar un esquema simple de identificación de rostros del vecino más cercano, dentro del espacio reducido *k*-dimensional.

1. Proyecte el conjunto de pruebas restando el vector medio multiplicado por el subespacio derivado *PCA Vk,* ver en la ecuación (4.26).
2. Calcule la distancia euclidiana entre cada cara en el conjunto de datos de entrenamiento contra el conjunto de prueba. El candidato con la menor distancia a la imagen de prueba se considera la punta de la nariz de un rostro.

$$\bar{X}_t^T = (X_t - \bar{X})^T\, V_k \tag{4.26}$$

Máquina de soporte vectorial (SVM)

La máquina de vectores de soporte (SVM) se considera un método de aprendizaje supervisado. Los conceptos básicos de SVM tienen su base en el trabajo sobre la teoría del aprendizaje estadístico introducido por Vapnik, Golowich y Smola (1996). La clasificación es una de las tantas tareas que se hacen con apoyo de los SVM y pertenecen a la categoría de clasificadores lineales, ya que producen separadores hiperplanos o lineales: espacio separable, cuasi-separable, transformado (Bouzalmat, Kharroubi y

Zarghili, 2014; Gumus *et al.,* 2010; Tello, Hernández-Ramírez y García-Sepúlveda, 2013). Mientras que la mayor parte de los métodos de aprendizaje minimizan los errores calculados por el modelo generado a partir de la muestra de entrenamiento, el sesgo asociado con el SVM radica en minimizar el riesgo estructural (Vapnik, Golowich y Smola, 1996).

Para la implementación de SVM, los vértices 3D han sido filtrados previamente como candidatos a punta de nariz. En este sentido, se debe seleccionar un hiperplano de separación equidistante de las muestras más cercanas a cada clase, en otras palabras, obtener un margen máximo a cada lado del hiperplano. Además, al definir el hiperplano, únicamente se consideran los datos de entrenamiento de cada clase que se encuentran en el borde de estos márgenes (vectores de soporte). Por lo tanto, los vectores de soporte son subconjuntos de observaciones de entrenamiento que se utilizan para respaldar una superficie de decisión de ubicación óptima (Vapnik, Golowich y Smola, 1996).

Entrenamos SVM con 200 histogramas de punta de nariz correctos y 200 histogramas con forma de punta de nariz, que en realidad eran ropa, cejas o mentón. Como se observa en el cuadro 4.3, los histogramas de punta de nariz correctos e incorrectos están etiquetados como 1 y -1, respectivamente. La implementación de SVM se llevó a cabo utilizando la caja de herramientas Matlab Support Vector Machine para clasificación.

Cuadro 4.3. Matriz de codificación de datos que ejemplifica la distribución del entrenamiento de SVM

	C_1	C_2	C_3	C_4	C_5	C_6	…	C_{256}	d_1
P_1	1	0	1	1	0	1	…	1	1
P_2	0	1	0	1	0	1	…	0	-1
P_3	1	0	1	1	0	1	…	1	1
…	…	…	…	…	…	…	…	…	…
P_{400}	1	0	1	1	0	1	…	1	1

Se utilizó una función de Matlab optimizada (*svmtrain*) para identificar los vectores de soporte S_i, los pesos $\propto_i$ y el sesgo para un kernel dado para clasificar un conjunto de pruebas x:

$$c = \sum_i \propto_i k(S_i, x) + b \qquad (4.27)$$

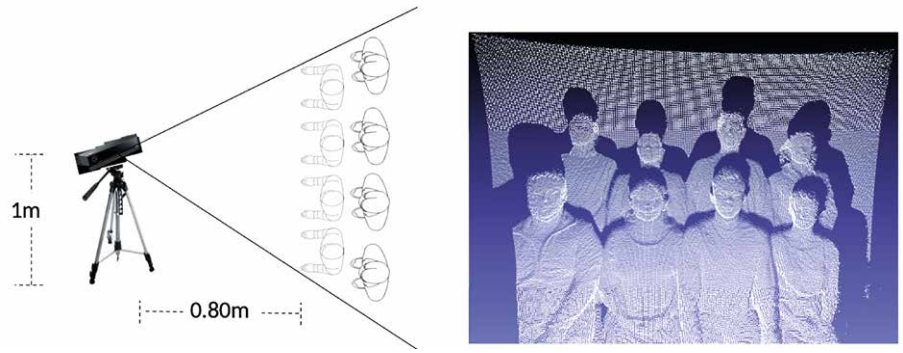
Se uso K–NN para calcular las distancias de nuestro conjunto de entrenamiento y los candidatos a la punta de la nariz, se selecciona el punto con la distancia más corta. Finalmente, de la punta de la nariz seleccionada se extrae la superficie facial de la imagen utilizando una radio de 8 cm.

Adquisición de imágenes experimentales

Como sabemos, las bases de datos 3D de última generación cuentan con solo una persona por imagen y fueron recopiladas en condiciones de iluminación controlada, con expresiones faciales neutrales y pose frontal (Senthilkumar y Gnanamurthy, 2014). Por lo tanto, no existen imágenes experimentales 3D para el propósito de esta investigación. Para experimentar la detección de rostros en imágenes 3D con una persona, se utilizaron dos bases de datos de última generación: Face Recognition Grand Challenge (Phillips *et al.,* 2005) y CurtinFaces (Li *et al.,* 2013), 4950 y 1437 imágenes 3D, respectivamente.

Para el experimento de detección de rostros en imágenes 3D con más de una persona, se hizo una sesión de captura utilizando la cámara Kinect One™, para recopilar imágenes 2D/3D variando el número y posición de hasta ocho sujetos en la escena.

Figura 4.5. Escenario de captura de una imagen de profundidad con ocho personas en la escena



Fuente: elaboración propia.

La figura 4.5 muestra cómo se colocaron las personas dentro del campo de visión de la cámara 3D. Nuestra recopilación de datos se limita a ocho personas por escena, ya que el campo de visión de la cámara utilizada no puede capturar más personas sin reducir el número de vértices por rostro. Este es un desafío bien conocido en el procesamiento de rostros 3D (Bowyer, Chang y Flynn, 2006).

Se recopilaron cuatro conjuntos de imágenes variando en cada escena el número y posición de ocho sujetos, lo que dio lugar a 28 posiciones diferentes y un total de 1 020 imágenes de personas sentadas en ocho posibles lugares. Para el entrenamiento y evaluación del rendimiento fue necesario recopilar la ubicación verdadera de la nariz. Al asumir que un rostro existe en una imagen si su nariz está presente, recopilamos la posición verdadera al nivel de vértice de la punta de la nariz, con identificación manual de la nariz como la parte más prominente de un rostro en una sesión 3D de Matlab.

Para implementar y evaluar el procedimiento experimental para detectar cualquier rostro presente en una imagen 3D, utilizando un conjunto de 1 020 imágenes de profundidad, definiendo diferentes conjuntos de entrenamiento y prueba como se indica en el cuadro 4.4.

Cuadro 4.4. Conjuntos de entrenamiento y prueba utilizando imágenes de profundidad de una a ocho personas diferentes

Conjunto	Número de imágenes	Número de rostros
Entrenamiento	100	200
Prueba	920	3 896
Total	1 020	4 096

Resultados experimentales

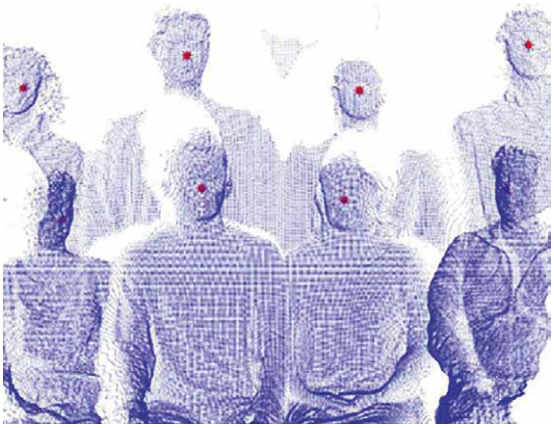
Al utilizar el conjunto de datos propuesto de rostros que contiene de una a ocho personas en cada escena, hemos evaluado experimentalmente el procedimiento propuesto de detección de rostros. En este caso, a partir de la clasificación de imágenes de giro resultante se encontró primero el vértice máximo local para cada clúster de candidatos. Luego, se calcularon errores de localización, como la distancia euclidiana entre la punta de la nariz localizada y su valor de verdad respectivo.

El cuadro 4.5 resume el rendimiento de la propuesta para la detección de rostros. Como podemos ver, se detectaron con éxito el 91% de los rostros en las imágenes experimentales de rostros, el 100% de los rostros en la base de datos FRGC y el 98% de los rostros en la base de datos CurtinFaces, dentro de un error de localización menor o igual a 16 mm. La figura 4.6 muestra una imagen con una detección correcta de ocho sujetos en escena, donde los puntos rojos indican la presencia de una nariz humana.

Cuadro 4.5. Resumen de la detección de rostros utilizando FRGC, CurtinFaces y el conjunto de imágenes propuestas conteniendo de uno a ocho sujetos a una precisión de 16mm

Base de datos	Rendimiento con PCA	Rendimiento con SVM
FRGC	98 %	100 %
CurtinFaces	97 %	98 %
Nuestra base de datos	85 %	91 %

Figura 4.6. Detección exitosa de ocho sujetos en la escena, los puntos rojos muestran las narices identificadas



Fuente: elaboración propia.

Conclusiones

En este capítulo se presentó un novedoso procedimiento experimental invariante a la posición para la detección automática de rostros en imágenes 3D, que consta de cuatro pasos: eliminación de regiones planas, análisis de curvatura, análisis y clasificación de imágenes de giro. Este enfoque de localización se evaluó utilizando nuestro conjunto de datos de imágenes 3D (que contiene de uno a ocho sujetos diferentes) y bases de datos de última generación (FRGC y CurtinFaces) que se recopilaron considerando solo una persona por imagen.

Al inicio de este trabajo, se investigaron experimentalmente tres enfoques clave en la literatura relacionada con nuestro objetivo final. El primer enfoque utiliza un análisis de curvatura para detectar el área facial más rígida donde se esperan las esquinas internas de los ojos y la punta de la nariz. El segundo enfoque corta horizontalmente una imagen de ingreso para localizar el área facial más probable que contiene la punta de la nariz. El tercer enfoque utiliza un procedimiento de segmentación para detectar un área elíptica donde se espera un rostro humano.

Para los tres conjuntos de datos experimentados (nuestra base de datos con imágenes con más de un sujeto, FRGC y CurtinFaces) hemos recopilado errores de ubicación calculando la distancia euclidiana desde la punta de la nariz seleccionada y su respectiva verdad sobre el terreno. Los mejores resultados experimentales: 454 utilizando un clasificador SVM, muestran que, dentro de un error de localización entre 0 y 16 mm, se ubican el 100%, el 98% y el 91% de las caras en cada imagen del FRGC, CurtinFaces y nuestro conjunto de datos.

A pesar de la baja resolución de Kinect One™ y el campo de visión limitado, nuestro enfoque de detección de rostros ha demostrado ser sólido para la localización de la punta de la nariz. Luego, nuestro enfoque supone la presencia de una cara al localizar la punta de su nariz. Por otro lado, nuestro enfoque de detección de rostros se experimentó utilizando umbrales preestablecidos que se calcularon experimentalmente para cada paso.

Es claro que la IA desempeña un papel crucial en la detección de rostros humanos en imágenes 3D. En este capítulo se investigó una vertiente de esa área, basada en técnicas de reconocimiento de patrones. Como trabajo futuro se investigarán formas alternas en que la IA se puede utilizar para este propósito, incluyendo: redes neuronales convolucionales (CNN), aprendizaje profundo, reforzando con sensores 3D y tecnologías de escaneo.

Referencias

Aghaei, M., Dimiccoli, M. y Radeva, P. (2016). Multi-face tracking by extended bag-of-tracklets in egocentric photo-streams. *Computer Vision and Image Understanding 149,* pp. 146-156.

Ariz, M., Bengoechea, J. J., Villanueva, A. y Cabeza, R. (2016). A novel 2d/3d database with automatic face annotation for head tracking and pose estimation. *Computer Vision and Image Understanding 148,* pp. 201-210.

Belahcene, M., Chouchane, A. y Mokhtari, N. (2014). 2D and 3D face recognition based on IPC detection and patch of interest regions. Connected Vehicles and Expo (ICCVE). *2014 International Conference on...,* pp. 627-628.

Besl, P. J. y Jain, R. C. (1986). Invariant surface characteristics for 3D object recognition in range images. *Computer Vision, Graphics, and Image Processing 33*(1), pp. 33-80.

Bhandarkar, S. M. y Luo, X. (2009). Integrated detection and tracking of multiple faces using particle filtering and optical flow-based elastic matching. *Computer Vision and Image Understanding 113*(6), pp. 708-725.

Bouzalmat, A., Kharroubi, J. y Zarghili, A. (2014). Comparative study of PCA, ICA, Ida using SVM classifier. *Journal of Emerging Technologies in Web Intelligence 6*(1), pp. 64-68.

Bowyer, K. W., Chang, K. y Flynn, P. (2006). A survey of approaches and challenges in 3d and multi-modal 3d + 2d face recognition. *Computer Vision and Image Understanding 101*(1), pp. 1-15.

Cai, D., He, X., Han, J. y Zhang, H. J. (2006). Orthogonal Laplacian faces for face recognition. *IEEE Transactions on Image Processing 15*(11), pp. 3608-3614.

Chang, K. I., Bowyer, K. W. y Flynn, P. J. (2006). Multiple nose region matching for 3d face recognition under varying facial expression. *IEEE Transactions on Pattern Analysis and Machine Intelligence 28*(10), pp. 1695-1700.

Chellappa, R., Wilson, C. y Sirohey, S. (1995). Human and machine recognition of faces: a survey. *Proceedings of the IEEE 83*(5), pp. 705-741.

Chen, W., Sun, T., Yang, X. y Wang, L. (2009). Face detection based on half face-template. Electronic Measurement Instruments, 2009. ICEMI'09. *9th International Conference on...,* pp. 4-58.

Colombo, A., Cusano, C. y Schettini, R. (2006). 3D face detection using curvature analysis. *Pattern Recognition 39*(3), pp. 444-455.

Creusot, C., Pears, N. y Austin, J. (2010). 3D face landmark labelling. Proceedings of the ACM workshop on 3D object retrieval. *ACM, As sociation for Computing Machinery,* pp. 27-32.

Gumus, E., Kilic, N., Sertbas, A. y Ucan, O. N. (2010). Eigenfaces and support vector machine approaches for hybrid face recognition. *Pattern recognition 8,* pp. 9.

Hjelmas, E. y Low, B. K. (2001). Face detection: A survey. *Computer Vision and Image Understanding 83*(3), pp. 236-274.

Hu, W., Hu, W. y Maybank, S. (2008). Adaboost-based algorithm for network intrusion detection. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 38*(2), pp. 577-583.

Hutton, T. J., Buxton, B. F. y Hammond, P. (2003). Automated registration of 3D faces using dense surface models. *Proc. British Machine Vision Conference*, pp. 439-448.

Johnson, A. E. y Hebert, M. (1998). Surface matching for object recognition in complex three-dimensional scenes. *Image and Vision Computing 16 (910),* pp. 635-651.

Kakadiaris, I. A., Toderici, G., Evangelopoulos, G., Passalis, G., Chu, D., Zhao, X., Shah, S. K. y Theoharis, T. (2016). 3D-2D face recognition with pose and illumination normalization. Computer Vision and Image Understanding. Universidad de Houston.

Krishna, K. y Murty, M. N. (1999). Genetic *k*-means algorithm. Systems, Man and Cybernetics. Part B: Cybernetics. *IEEE Transactions on 29*(3), pp. 433-439.

Kumar, A., Datta, M. y Kumar, P. B. (2015). Face Detection and Recognition: Theory and Practice. Chapman and Hall/CRC.

Li, B. Y. L., Mian, A. S., Liu, W. y Krishna, A. (2013). Using kinect for face recognition under varying poses, expressions, illumination and disguise. Applications of Computer Vision (WACV), *2013 IEEE Workshop on,* pp.186-192.

Maes, C., Fabry, T., Keustermans, J., Smeets, D., Suetens, P. y Vandermeulen, D. (2010). Feature detection on 3D face surfaces for pose normalisation and recognition. Biometrics: Theory Applications and Systems (BTAS). *2010 Fourth IEEE International Conference on,* pp. 1-6.

Mian, A., Bennamoun, M. y Owens, R. (2006). Automatic 3D face detection, normalization and recognition. *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, pp. 735-742.

Nair, P. y Cavallaro, A. (2009). 3-D face detection, landmark localization, and registration using a point distribution model. *Multimedia, IEEE Transactions on 11*(4), pp. 611-623.

Nanni, L., Lumini, A., Dominio, F. y Zanuttigh, P. (2014). Effective and precise face detection based on color and depth data. *Applied Computing and Informatics 10*(1-2), pp. 1-13.

Paduano, J., Romero, M. y Muñoz, V. (2015). Toward face detection in 3D data. *International Conference on Image Processing, Computer Vision, and Patter recognition.* Vol. 15, pp. 473-479.

Paduano, J., Romero, M. y Valdovinos, R. (2016). Face detection in 3D images with more than one person. *International Conference on Image Processing, Computer Vision, & Patter recognition,* pp. 287-293.

Pan, H., Zhu, Y. y Xia, L. (2013). Efficient and accurate face detection using heterogeneous feature descriptors and feature selection. *Computer Vision and Image Understanding 117*(1), pp. 12-28.

Pears, N., Liu, Y. y Bunting, P. (2012). 3D Imaging. Analysis and Applications. Vol. 1. Springer-Verlag.

Peng, X. y Bennamoun, M. (2011). A training-free nose tip detection method from face range images. *Pattern Recognition 44*(3), pp. 544-558.

Phillips, P. J., Flynn, P. J., Scruggs, T., Bowyer, K. W., Chang, J., Homan, K., Marques, J., Min, J. y Worek, W. (2005). Overview of the face recognition grand challenge. Computer Vision and Pattern Recognition, 2005. CVPR 2005. *IEEE Computer Society Conference on 1,* pp. 947-954.

Segundo, M. P., Queirolo, C., Bellon, O. R. P. y Silva, L. (2007). Automatic 3D facial segmentation and landmark detection. In: Image Analysis and Processing, 2007. ICIAP 2007. *14th International Conference on,* pp. 431-436.

Senthilkumar, R. y Gnanamurthy, R. K. (2014). A detailed survey on 2D and 3D still face and face video databases part I. *Communications and Signal Processing (ICCSP), 2014 International Conference on,* pp. 1405-1409.

Smeulders, A. W. M., Chu, D. M., Cucchiara, R., Calderara, S., Dehghan, A. y Shah, M. (2014). Visual tracking: An experimental survey. Pattern Analysis and Machine Intelligence, *IEEE Transactions on 36*(7), pp. 1442-1468.

Soldera, J., Behaine, C. A. R. y Scharcanski, J. (2015). Customized orthogonal locality preserving projections with soft-margin maximization for face recognition. IEEE *Transactions on Instrumentation and Measurement 64*(9), pp. 2417-2426.

Subburaman, V. B. y Marcel, S. (2013). Alternative search techniques for face detection using location estimation and binary features. *Computer Vision and Image Understanding 117*(5), pp. 551-570.

Tello, J. C. C., Hernández-Ramírez, D., García-Sepúlveda, C. A. (2013). Support vector machine algorithms in the search of kir gene associations with disease. *Computers in Biology and Medicine 43*(12), pp. 2053-2062.

Trucco, E. y Verri, A. (1998). *Introductory Techniques for 3-D Computer Vision.* Prentice-Hall.

Valstar, M. F., Mehu, M., Jiang, B., Pantic, M. y Scherer, K. (2012). Meta-analysis of the first facial expression recognition challenge. Systems, Man, and Cybernetics, Part B: Cybernetics. *IEEE Transactions on 42*(4), pp. 966-979.

Vapnik, V., Golowich, S. E. y Smola, A. (1996). Support vector method for function approximation, regression estimation, and signal processing. *Advances in Neural Information Processing Systems 9.* Citeseer, 1996.

Wagner, A., Wright, J., Ganesh, A., Zhou, Z., Mobahi, H. y Ma, Y. (2012). Toward a practical face recognition system: Robust alignment and illumination by sparse representation.

Pattern Analysis and Machine Intelligence. *IEEE Transactions on 34*(2), pp. 372-386.

Wang, P. y Ji, Q. (2007). Multi-view face and eye detection using discriminant features. *Computer Vision and Image Understanding 105*(2), pp. 99-111.

Wohlhart, P. y Lepetit, V. (2015). Learning descriptors for object recognition and 3D pose estimation. *Institute for Computer Vision and Graphics.* Graz University of Technology.

Yang, M. y Ahuja, N. (2001). Face Detection and Gesture Recognition for Human-Computer Interaction, *The International Series in Video Computing 1.*

Yang, M., Feng, Z., Shiu, S. C. K. y Zhang, L. (2014). Fast and robust face recognition via coding residual map learning based adaptive masking. *Pattern Recognition 47*(2), pp. 535-543.

Yang, M., Kriegman, D. y Ahuja, N. (2002). Detecting faces in images: a survey. Pattern Analysis and Machine Intelligence. *IEEE Transactions on 24*(1), pp. 34-58.

Yin, S., Ouyang, P., Dai, X., Liu, L. y Wei, S. (2015). An adaboost-based face detection system using parallel configurable architecture with optimized computation. *IEEE Systems Journal PP* (99), pp. 1-12.

Zafeiriou, S., Zhang, C. y Zhang, Z. (2015). A survey on face detection in the wild: Past, present and future. *Computer Vision and Image Understanding 138,* pp. 1-24.

Zhang, J., Zhang, X.-D. y Ha, S.-W. (2008). A novel approach using PCA and SVM for face detection. Natural Computation, 2008. ICNC'08. *Fourth International Conference on.* Vol. 3, pp. 29-33.

CIFRADO DE META-APRENDIZAJE EN REDES NEURONALES ARTIFICIALES

Graciela García Rueda
Facultad de Ingeniería, UAEMEX

Rosa María Valdovinos Rosas
Facultad de Ingeniería, UAEMEX

Javier Salas García
Facultad de Ingeniería, UAEMEX

Resumen

Las redes neuronales artificiales son modelos de aprendizaje automático para el procesamiento de grandes volúmenes de datos de manera eficiente y amplia aplicación. No obstante, en muchas ocasiones la información que se desea analizar y procesar contiene datos sensibles, por lo que la privacidad del meta-aprendizaje obtenido por estos modelos es de suma importancia. Esto lleva al uso de estrategias que ayuden a conservar la privacidad y confidencialidad en su proceso de análisis, por lo que se han hecho diferentes investigaciones utilizando métodos de cifrado dentro de algoritmos de aprendizaje automático. Derivado de lo anterior, en este capítulo se presenta la implementación de un algoritmo para cifrar el meta-aprendizaje de un perceptrón multicapa, con el propósito de validar la pertinencia de uso sin perder precisión en su desempeño al momento de clasificar nuevos casos.

Introducción

En los últimos años, la IA ha tenido una aplicación cada vez mayor en diferentes áreas, al proporcionar diferentes algoritmos, métodos y técnicas que pueden ser incluidos en diferentes aplicaciones que brindan solución a problemas específicos en diferentes

sectores, como la seguridad informática, la educativa, entre otros (Morales Peña *et al.,* 2010).

Una rama de la IA es el *machine learning* (ML), también conocido como aprendizaje automático o aprendizaje de máquina, en el que se crean sistemas que aprenden de forma automática o semi-automática para identificar patrones complejos en grandes volúmenes de datos. Estas técnicas se han implementado de forma satisfactoria en distintas áreas, que van desde la visión computacional, el reconocimiento de patrones en el sector médico; así como en la economía, la industria del entretenimiento, entre otras (El Naqa y Murphy, 2015).

En su parte más básica, los modelos de ML han cambiado de forma progresiva de un "estilo de entrenamiento cerrado" a un sistema inteligente más complejo. Al mismo tiempo, el entorno de nube abierta proporciona fuentes de datos y recursos informáticos adecuados para esta disciplina, por lo que el procesamiento de información a gran escala se puede hacer de forma eficaz (Sung *et al.,* 2015) De los algoritmos de ML más populares se pueden mencionar las redes neuronales artificiales (RNA), gracias a su potencia para la generalización del conocimiento y la velocidad que tienen al momento de realizar la clasificación de nuevos casos, lo que ha generado una creciente demanda de modelos en la nube, casos de uso en los que los datos y el modelo son propiedad de diferentes partes. Sin embargo, esto crea una serie de problemas de privacidad, por lo que la seguridad en los datos y el mismo modelo de RNA se ha convertido en un aspecto fundamental por atender (Novas Otero, 2018; Li *et al.,* 2017).

Un caso específico donde se utilizan RNA con datos confidenciales, en el que se requiere que tanto los datos como los modelos permanezcan privados, es el relacionado a la predicción de la probabilidad de readmisión de un paciente dentro de los próximos 30 días en un hospital, con el fin de mejorar la calidad de la atención y reducir los costos hospitalarios. Debido a los requisitos éticos y legales relacionados con la confidencialidad de la información del paciente, es posible que el hospital tenga prohibido utilizar dicho servicio; por lo tanto, el hospital cifra la información privada y la envía para realizar predicciones mediante un servicio en la nube que ofrezca la confidencialidad y privacidad de dichos datos (Gilad-Bachrach *et al.,* 2016).

Anteriormente, la protección de la información se fundamentaba, sobre todo, en el uso de distintos dispositivos para proteger las instalaciones en las que se alojaban los

servidores electrónicos y/o la información. Sin embargo, en la actualidad las medidas anteriores son obsoletas debido a la digitalización de la información y los formatos de almacenamiento y su acceso por medio de internet. Esto ha provocado que las personas puedan tener acceso de forma rápida y desde cualquier parte del mundo (Cuenca Guachamin, 2019).

Con base en lo anterior, una forma de proteger la privacidad al utilizar estas técnicas es mediante el uso de métodos de cifrado. Esto se define como el proceso de convertir mensajes, información o datos en un formato ilegible para cualquier persona, excepto el destinatario previsto (Yi, Paulet y Bertino, 2014). Para ello, los métodos de cifrado deben cumplir con algunos requisitos (Chase *et al.,* 2017): la privacidad o confidencialidad, la integridad y la autenticación, con el objetivo de garantizar una mejor seguridad. Existen diferentes tipos de cifrado, como el cifrado homomórfico, que permite hacer cálculos u operaciones matemáticas complejos en datos cifrados sin descifrarlos en todo el proceso de análisis; el cifrado simétrico, que mediante una sola clave se cifra y se descifra, y el cifrado asimétrico, en el que se hace uso de dos claves para el mismo propósito (Pousa, 2011).

De los métodos existentes, el cifrado homomórfico es uno de los más prometedores y es en el que se centra la investigación presentada en este capítulo.

En la literatura se pueden encontrar varios estudios que consideran el cifrado al utilizar algoritmos de ML. Tilen Marc *et al.* (2019), utilizan un cifrado funcional, para construir modelos eficientes de ML con privacidad mejorada y proporcionan una implementación de tres servicios de predicción que se pueden aplicar en los datos cifrados, mediante una red neuronal artificial.

Por otro lado, Li *et al.* (2017) presentan un esquema de cifrado totalmente homomórfico de múltiples claves y un esquema de cifrado basado en una estructura híbrida combinando el mecanismo de doble descifrado y el cifrado totalmente homomórfico, y agregan estos esquemas a un modelo de red neuronal y CryptoML.

Por su parte Pedro Novas Otero (2018) utiliza redes neuronales al criptoanálisis en el algoritmo Advanced Encryption Standard. Una red neuronal es entrenada con datos formados por pares de texto plano y sus correspondientes textos cifrados para, posteriormente, a partir de un texto cifrado predecir el texto en plano correspondiente. Como resultados del estudio se evidenció la capacidad de aprendizaje de las redes neuronales probadas en versiones limitadas del algoritmo. Sin embargo, en las versiones

totales del algoritmo las pruebas muestran una aleatoriedad en las predicciones, lo que hace que exista un error en la precisión y no sea realmente confiable.

Como puede verse, los trabajos revisados se centran en el cifrado de las capas de la red y de los datos utilizados, por lo que en este estudio se presenta el cifrado del vector de pesos de la red neuronal, como resultado del proceso de aprendizaje del perceptrón multicapa, mismo que se utilizará en la etapa de clasificación. Los modelos cifrados son validados en 10 conjuntos de datos, cuyos resultados muestran la pertinencia de su aplicación en términos de clasificación y costo computacional.

Cifrado de Paillier

Un método de cifrado asegura que los datos permanezcan confidenciales, por lo que es un elemento fundamental de la seguridad y es la forma más simple e importante de impedir que alguien robe o lea la información con fines malintencionados (Prieto Rodríguez, 2015).

Uno de los métodos de cifrado homomórfico más popular es el cifrado de Paillier, un método de cifrado asimétrico probabilístico con propiedades homomórficas que permiten la suma y multiplicación de sus textos cifrados. En su funcionamiento, consta de al menos tres procesos: generación de claves, cifrado y descifrado. El primero convierte textos sin formato en cifrados y el descifrado tiene el efecto inverso. La seguridad del método se basa en la factorización de un número N que ha sido construido a partir del producto de dos primos (Rodríguez Henríquez, 2009).

Para realizar el cifrado de Paillier utiliza dos claves públicas g y n, compuestas por los números primos p y q distintos entre ellos (Aravena Bravo *et al.*, 2018), y dos claves privadas μ y λ, donde $\lambda = lcm((p - 1), (q - 1))$. La generación de las claves (pública y privada), se hace de la siguiente manera:

1. Se eligen dos números primos largos, p y q, los cuales deben cumplir que el máximo común divisor (*gcd*) sea igual a uno, tal como se observa en la ecuación 5.1:

$$gcd(pq, ((p-1)(q-1))) = 1 \qquad (5.1)$$

Esta propiedad está asegurada si ambos primos tienen la misma longitud.

2. Se calcula un módulo *n* equivalente a $n = pq$, luego se calcula la función de Carmichael (un entero positivo *n*, denotada $\lambda(n)$, se define como el menor entero *m* tal, que cumple: $a^m \equiv (mod\ n)$), utilizando la ecuación 5.2:

$$\lambda = \frac{(p-1)(q-1)}{gcd((p-1)(q-1))} \tag{5.2}$$

Esta propiedad está asegurada si ambos primos tienen la misma longitud.

3. Se selecciona un número aleatorio *g* donde $g \in Z_{n^2}^*$

$$gcd = \frac{g^{\lambda} mod\ n^2 - 1}{1} = 1 \tag{5.3}$$

4. Por último, se aplica el inverso multiplicativo con la ecuación 5.4:

$$\mu = (L(g^{\lambda} mod\ n^2))^{-1} mod\ n \tag{5.4}$$

Donde: *L* es: $L(u) = (u - 1) / n$

5. Como resultado se obtienen las claves:

- Clave pública (cifrado): *(g, n)*
- Clave privada (descifrado): *(μ, λ)*

Al utilizar las claves generadas, el proceso de cifrado de un mensaje *M* se lleva acabo de la forma siguiente:

1. Sea *M* un mensaje para ser cifrado, tal que $M \in \mathbb{Z}_n$
2. Seleccionar al azar *r*, donde $0 \leq r \leq n$ y $r \in Z_n^*$; es decir, asegurar la igualdad $gcd(r, n)=1$

3. Calcular el texto cifrado como lo indica la ecuación 5.5:

$$c = g^m * r^n mod\ n^2 \tag{5.5}$$

Mientras que, para el proceso de descifrado, se hace lo siguiente:

1. Dejar que c sea el texto cifrado a descifrar, donde $c \in Z^*_{n^2}$

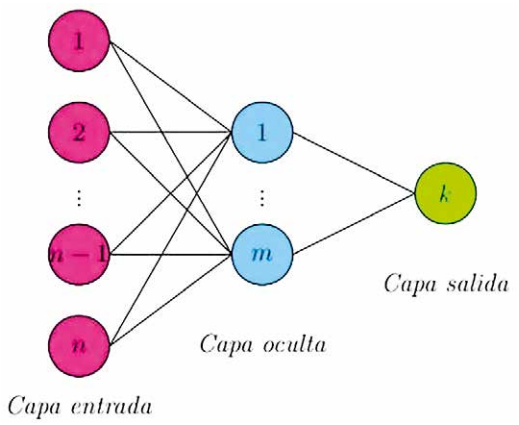
2. Calcular el mensaje de texto sin formato como se indica en la ecuación 5.6:

$$m = L(c^\lambda\ mod\ n^2) * \mu\ mod\ n \tag{5.6}$$

Perceptrón multicapa

El perceptrón multicapa es una red neuronal constituida por tres o más capas (Rodríguez Ponce, 2014) y una estructura como la mostrada en la figura 5.1.

Figura 5.1. Perceptrón multicapa



Fuente: Rodríguez Ponce, 2014.

Sean x_i, las entradas de la red; y_j, las salidas de la capa oculta; z_k, las salidas de la capa final; $w_{i\ j}$, los pesos de la capa oculta y θ_j, sus umbrales; w', los pesos de la capa de

salida, y θ', sus umbrales, para todo $i = 1, ..., n$, $j = 1, ..., q$ y para todo $k = 1, ..., m$. Matemáticamente se representa mediante la siguiente fórmula, donde f es la función de activación que se muestra en la ecuación 5.7:

$$z_k = \sum_{j=1}^{q} w'_{kj} y_j - \theta'_k = \sum_{j=1}^{q} w'_{kj} f\left(\sum_{i=1}^{n} w_{ji}\, x_i - \theta_j\ \right) - \theta'_k \qquad (5.6)$$

El aprendizaje de un perceptrón multicapa tiene por objetivo la minimización del error que mide la diferencia entre la salida z obtenida por la red y la salida deseada t.

El proceso de aprendizaje más utilizado es el *Backpropagation,* que se explica en el algoritmo 2, donde en la línea 3 se aplica la selección para cada ejemplo del conjunto de datos de entrenamiento, y este se itera hasta que el error baje de un umbral. La siguiente fase es la propagación hacia delante, en la que se hace el cálculo de la salida de la red (*yk*) como se muestra en la línea 5, para después hacer el cálculo de los δ en la última capa.

Figura 5.2. Algoritmo 2 *Backpropagation*

Algorithm 1 *Backpropagation*

Require: red conjunto, η
Ensure: Red
1: $\{W_{ij}\} \leftarrow$ Inicializar;
2: **while** $\neg\, Convergencia(red)$ **do**
3: $e \leftarrow SeleccionConjunto(conjunto)$;
4: $\{y_k\} \leftarrow Forward(e)$;
5: $\{D_k\} \leftarrow SalidaDeseada(e)$;
6: **for** $n_k \in Capa(red, k)$ **do**
7: $\delta_k = (d_k - y_k) f'(net_k)$;
8: **end for**
9: **for** $j = k - 1, j <= 1$ **do**
10: **for** $n_j \in Capa(red, j)$ **do**
11: $\delta_j = f'(net_j) \Sigma_{j+i} \delta_{k+1} w_{j(j+1)}$;
12: **end for**
13: **end for**
14: **for** $j = k, j <= 1$ **do**
15: $w_{(j-1)j} = w_{(j-1)j} + \eta \delta_j y_{(j-1)}$;
16: **end for**
17: $red \leftarrow ActualizarRed(\{w_{ij}\})$;
18: **end while**

Una vez que la red ha finalizado su proceso de entrenamiento, el meta-aprendizaje se guarda en el vector de pesos *w,* que es utilizado en la etapa de clasificación.

Es importante que el meta-aprendizaje de la red se mantenga íntegro y no se corrompa a fin de asegurar su desempeño, además de ser alterado, el reconocimiento que la red pueda hacer también cambiaría. Esto último es lo que justifica la implementación de una etapa de cifrado, a fin de garantizar y mantener consistente el conocimiento de la red.

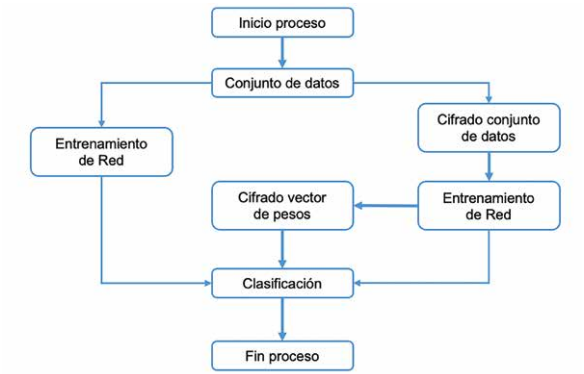
Resultados

El objetivo de la experimentación presentado en este capítulo se centró en validar la pertinencia de cifrar el meta-aprendizaje de un perceptrón multicapa para brindar mayor seguridad al conocimiento aprendido de la red y determinar que, derivado del proceso de descifrado, no se haya perdido información en el proceso de clasificación de nuevos patrones.

En la experimentación se diseñaron tres escenarios experimentales que se muestran a continuación y se reflejan en el diagrama de la figura 5.3.

1. Datos fuente y modelo de red sin cifrar. Este escenario es considerado como valor de referencia.
2. Datos fuentes sin cifrar y modelo de red cifrado. Para dicho escenario se utilizó el modelo de cifrado Paillier mencionado en una sección anterior para cifrar el vector de pesos resultante del entrenamiento, dejando los datos fuente sin cifrar.
3. Datos fuente cifrados y modelo de red sin cifrar. Es decir, cifrar cada uno de los datos fuentes utilizados para entrenar el modelo de red.

Figura 5.3. Diagrama de escenarios



Fuente: elaboración propia.

Los conjuntos de datos utilizados fueron obtenidos de *UC Irvine Machine Learning Repository*. En todos los conjuntos de datos se utilizó Validación cruzada con 10 repeticiones y una distribución de 80% para entrenamiento y 20% para fines de prueba.

La configuración del modelo de red utilizado es Capas ocultas 5, Capas de salida 1, Épocas de 2500, la inicialización de pesos se hizo de forma aleatoria, un *batch_size* de 5, una activación tipo *relu* para las capas de salida y entrada, mientras que para la capa de salida se ocupó una activación tipo *sigmoid*, una función de pérdida tipo *binary_crossentropy* y una métrica *accuracy*.

Los resultados obtenidos se presentan desde dos perspectivas: la primera, que es la precisión de la red al clasificar nuevos casos, y la segunda, el tiempo requerido por el entrenamiento de la red.

Precisión de los modelos de red

Para analizar el rendimiento, en términos de precisión, de los modelos de red utilizados, el cuadro 5.1 muestra los resultados obtenidos de los tres escenarios experimentales. La segunda columna incluye los resultados obtenidos con el modelo de red sin cifrar, que son considerados como el valor de referencia. La tercera y cuarta

columna muestran los resultados considerando el método de cifrado en el escenario experimental 2 y 3.

Considerando los valores de referencia del modelo sin cifrar es posible identificar un desempeño semejante entre la utilización de ambas estrategias, cifrado del modelo *Vs* cifrado de los datos. Esto último es altamente deseable, ya que uno de los efectos adversos es la pérdida de precisión al incluir una estrategia de cifrado al modelo de red, lo que podría indicar pérdida de información o alteración del meta-aprendizaje de la red ya entrenada, para el caso en el que se cifran los datos.

Tiempo de ejecución

Incluir una estrategia de cifrado al modelo de red forzosamente implica incrementar el costo computacional requerido por el modelo. No obstante, lo importante en este punto es determinar el costo-beneficio que la propuesta tiene.

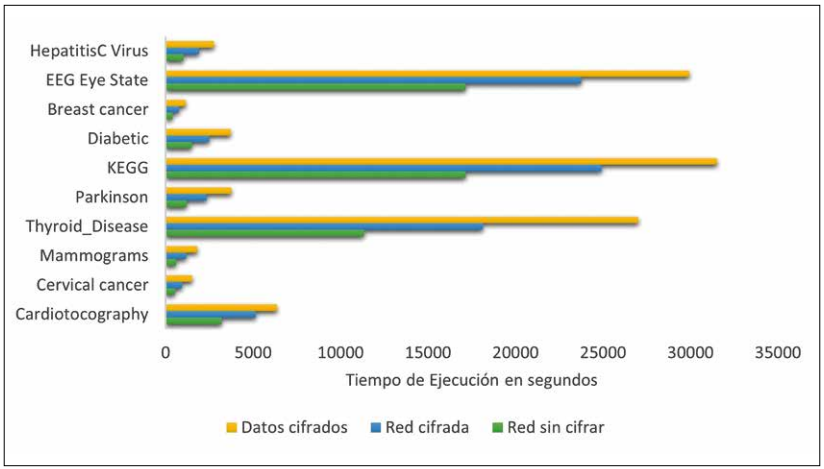
Al respecto, la gráfica de la figura 5.4 muestra los tiempos requeridos por los tres escenarios experimentales analizados en este capítulo. En la gráfica, el eje *y* marca el tiempo de ejecución, en segundos, requerido al procesar cada uno de los conjuntos de datos utilizados.

Como se puede observar, el tiempo de ejecución considerando el modelo sin cifrar es el menor de los tres escenarios. Al revisar el tiempo requerido por los dos escenarios que consideran cifrado en su proceso, es posible ver que el tiempo de ejecución aumentó, no obstante el tiempo que tardó el método en cifrar el vector de pesos no excede el doble del tiempo del modelo sin cifrar.

Cuadro 5.1. Resultados de exactitud con los tres escenarios experimentales

Conjunto datos	Red no cifrada	Red cifrada	Datos cifrados
Cardiotocography	95.0 %	94.2 %	94.7 %
Cervical cancer (risk_factors)	95.7 %	95.0 %	94.3 %
Mammographic Masses	98.0 %	98.0 %	98.0 %
Thyroid_Disease	97.7 %	96.0 %	97.3 %
Parkinson Multiple Sound	97.0 %	97.8 %	96.4 %
KEGG Metabolic Reaction	98.0 %	98.1 %	97.0 %
Diabetic Retinopathy Debrecen	99.0 %	98.3 %	98.0 %
Breast cancer Wisconsin	90.4 %	92.0 %	90.1 %
EEG Eye State	95.08 %	95.0 %	95.08 %
Hepatitis C Virus (HCV)	96.0 %	96.0 %	95.7 %

Figura 5.4. Tiempo de ejecución requerido por los modelos de red



Fuente: elaboración propia.

Respecto al tercer escenario es de esperarse que sea mayor que cifrar solamente el meta-aprendizaje, ya que el cifrado de datos, es exponencial a la cantidad de instancias de cada conjunto de datos.

Conclusiones

Al procesar información, mantener la privacidad, seguridad de los datos utilizados es fundamental. En aprendizaje automático, las redes neuronales artificiales son de los modelos más socorridos para hacer tareas tanto de descripción, como de predicción. Pese al rendimiento tan favorable que estos modelos llegan a tener, es fundamental incluir alguna estrategia orientada a salvaguardar tanto el meta-aprendizaje, como la integridad de los datos fuente.

La investigación hecha en este capítulo se centró en analizar la pertinencia de utilizar el método de Paillier para cifrar, ya sea el meta-aprendizaje de un perceptrón multicapa con aprendizaje *Backpropagation*, o los datos utilizados para su entrenamiento. Para este fin, la experimentación se orientó al análisis de dos variables, la precisión de los modelos al clasificar nuevos casos y el tiempo de ejecución requerido.

Los resultados obtenidos permitieron observar que 1) utilizar el cifrado de los datos o del meta-aprendizaje de la red aseguran los índices de precisión semejantes a los obtenidos cuando el modelo de red no es cifrado; 2) al analizar las dos estrategias para cifrar el modelo de red tampoco se observa diferencia significativa en términos de exactitud; 3) el costo computacional requerido por el cifrado de los datos, en algunos casos, iguala o supera el doble del tiempo requerido por el modelo sin cifrar.

Derivado de lo anterior, con la investigación aquí presentada, se pone de manifiesto la pertinencia de cifrar el meta-aprendizaje del modelo de red en lugar de cifrar los datos de entrada, ya que se obtienen índices de precisión y se requiere menos del 40% adicional de tiempo al requerido por el modelo sin cifrar.

Como líneas abiertas de estudio se plantea implementar estrategias orientadas a disminuir el tiempo en el cifrado del meta-aprendizaje de la red, además de extender la propuesta a modelos convolucionales de aprendizaje profundo y la utilización de métodos estocásticos de cifrado.

Referencias

Aravena Bravo, M. O. *et al.* (2018). Selección e implementación de librería Java de algoritmo para E-Voting. [Tesis doctoral.] Universidad del Bío-Bío.

Chase, M. *et al.* (2017). Security of homomorphic encryption. HomomorphicEncrption. Org. Redmond WA, Tech. Rep.

Cuenca Guachamin, W. R. (2019). *Técnicas de machine learning aplicadas a la seguridad.* [Tesis de maestría.] Universitat Oberta de Catalunya [UOC].

El Naqa, I. y Murphy, M. J. (2015). What is machine learning? *Machine Learning in Radiation Oncology,* pp. 3-11.

Gilad-Bachrach, R. *et al.* (2016). Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy. *International Conference on Machine Learning*. PMLR, pp. 201-210.

Li, P. *et al.* (2017). Multi-key privacy-preserving deep learning in cloud computing. *Future Generation Computer Systems 74,* pp. 76-85.

Morales Peña, R. *et al.* (2010). La inteligencia artificial en la actualidad. *Revista Tecnológica 3*(3), pp. 17-20.

Novas Otero, P. (2018). Redes neuronales aplicadas al criptoanálisis del Advanced Encryption Standard. *Universitat Oberta de Catalunya*.

Pousa, A. (2011). *Algoritmo de cifrado simétrico AES*. [Tesis de maestría.] Universidad Nacional de La Plata.

Prieto Rodríguez, J. D. (2015). Algoritmo de generación de llaves de cifrado basado en biometría facial. *INVENTUM 10.19,* pp. 41-51.

Rodríguez Henríquez, L. M. X. (2009). *Esquema de cifrado para la ejecución de consultas en bases de datos cifradas*. [Tesis doctoral.] Instituto Politécnico Nacional.

Rodríguez Ponce, H. U. (2014). *Perceptrón multicapa para reconocimiento de objetos sobre planos*. [Tesis doctoral.] Universidad del Cauca.

Sun, G. *et al.* (2015). MLaaS: a cloud-based system for delivering adaptive micro learning in mobile MOOC learning. *IEEE Transactions on Services Computing 11.2,* pp. 292-305.

Tilen Marc *et al.* (2019). Privacy-enhanced machine learning with functional encryption. *European Symposium on Research in Computer Security,* pp. 3-21.

Yi, X., Paulet, R. y Bertino, E. (2014). Homomorphic encryption. *Homomorphic Encryption and Applications,* pp. 27-46.

APRENDIZAJE AUTOMÁTICO: TENDENCIAS Y DESAFÍOS ACTUALES

Héctor Alejandro Montes-Venegas
Facultad de Ingeniería, UAEMEX

Resumen

El aprendizaje automático (AA), área relevante de la IA, ha transitado en años recientes de un campo de estudio para académicos de las Ciencias Computacionales a una fuerza de creciente interés en múltiples áreas, como la medicina, la investigación científica aplicada, e incluso, en la vida cotidiana en el planeta. Su habilidad para extraer patrones de datos, hacer predicciones, tomar decisiones o sintetizar y generar conocimiento, ha logrado avances significativos y también ha provocado diversos temores. Sin embargo, al igual que sucede con otras áreas de rápido crecimiento, también ha generado diversas tendencias, expectativas y retos importantes.

Este capítulo explora los fundamentos del aprendizaje automático, sus diversas aplicaciones, y los numerosos retos que investigadores y usuarios están enfrentando, así como el considerable impacto que tiene y que seguirá teniendo en nuestras vidas y en nuestro entorno.

Introducción

El aprendizaje automático (AA) o *machine learning* (nombre con el que originalmente fue concebido en 1959 por Arthur Samuel, pionero en juegos por computadora) (Samuel, 2010), es una rama de la IA que rápidamente ha causado que gobiernos, industrias y la sociedad, en general, concentren parte de su atención y sus recursos en ella (Ngai, Lui y Lee, 2022).

A finales de la década de los noventa, Tom Mitchell, conocido por sus contribuciones al AA (Mitchell, 1997a), se cuestionaba: "¿Realmente funciona el aprendizaje automático?" (Mitchel, 1997b). Su respuesta categórica fue "Sí", y

relataba cómo en el transcurso de esa década, el AA "había evolucionado de un campo de demostraciones de laboratorio a un campo de valor comercial significativo".

El potencial transformador que recientemente ha mostrado en diversas áreas como el comercio, la agricultura, la medicina, la investigación científica e, incluso, en el diario vivir, hacen aún más válida, un cuarto de siglo más tarde, la respuesta tajante de Mitchell. Esto se ve reflejado en la larga lista de aplicaciones logradas con IA, en general y particularmente, con el AA. Estos logros han tenido una extensa difusión en medios científicos e informativos de todo tipo. Sin embargo, el AA también ha generado un sinnúmero de expectativas, retos y hasta riesgos importantes. Es muy probable que muchos de ellos hayan surgido a raíz del hecho de que, finalmente, hemos podido hacer que una máquina, además de procesar información y proveernos de datos útiles, tenga la capacidad de entablar una conversación (*Natural Machine Intelligence,* 2023) y retar al "juego de la imitación" propuesto por Alan Turing como un mecanismo para identificar inteligencia (Turing, 1950).

Antes de discutir una serie de retos y expectativas del AA, abordaremos brevemente algunos de sus fundamentos y áreas de aplicación relevantes.

Fundamentos del aprendizaje automático

Casi desde que las computadoras fueron inventadas se ha cuestionado si son capaces de aprender como una forma de exhibir inteligencia. El aprendizaje humano es el proceso cognitivo de adquirir y modificar el conocimiento como resultado de la experiencia (Schemmer *et al.* 2023). Este mismo proceso es el que se emula en el AA a través del estudio y desarrollo de algoritmos capaces de procesar datos y adquirir conocimiento de ellos para luego ampliar y/o generalizar su aprendizaje sobre otros datos que el mismo algoritmo no había examinado previamente.

Como consecuencia de este proceso de adquisición de conocimiento, resulta evidente que, entre mayor sea la calidad de los datos que el AA tenga disponibles, mayor será la precisión que se obtendrá en la aplicación deseada.

En general, el AA implica la extracción de conocimiento de un cúmulo de datos y, para lograrlo, los algoritmos de AA construyen un modelo o representación basado en

las relaciones intrínsecas de los datos analizados. Como resultado, varios paradigmas se han establecido para producir soluciones a diversos problemas de aplicación práctica.

Paradigmas del aprendizaje automático

Es común dividir las diferentes técnicas de AA en tres categorías, cada una de ellas se ha convertido en un paradigma de aprendizaje. Y aunque los diferentes algoritmos en cada categoría tienen ventajas y desventajas, ninguno de ellos ha demostrado funcionar para todos los problemas tratados (Gaur *et al.,* 2019).

Aprendizaje supervisado

En el aprendizaje supervisado, los algoritmos se entrenan con un conjunto de datos etiquetados o anotados, donde los datos de entrada se asocian con etiquetas de salida correspondientes; esto es, los objetos de entrada y un valor deseado de salida se utilizan para entrenar un modelo de aprendizaje. Una vez entrenado, el modelo se utiliza para procesar otros datos no vistos durante su fase de entrenamiento. De esta forma, se espera que su desempeño sea adecuado y que, en consecuencia, el modelo haya aprendido a mapear las entradas a las salidas, lo que le permite hacer predicciones sobre nuevos datos no vistos previamente.

Por ejemplo, supongamos que necesitamos un modelo para identificar imágenes de un objeto de interés dado. Para esto, se deben tener disponibles una cantidad considerable de datos (imágenes) que contienen el objeto de interés aparejado con su correspondiente etiqueta. El algoritmo utilizado para entrenar el modelo analiza estos pares de información para luego inferir cuál sería la etiqueta adecuada cuando se le pida hacer una predicción con imágenes nuevas.

Los algoritmos modernos del AA supervisado hacen posible que sea menos complejo para organizaciones de diversas áreas crear modelos muy complejos capaces de hacer predicciones con altos grados de precisión. Esto ha causado que se utilicen en diferentes áreas como mercadotecnia, servicios financieros, salud, producción de energía, entretenimiento, seguridad, entre muchos otros.

Aprendizaje no supervisado

El aprendizaje no supervisado implica trabajar con datos no etiquetados. Esta forma de aprendizaje explora la estructura inherente dentro de los datos, a menudo descubriendo patrones, agrupaciones o relaciones, sin recibir la guía explícita dada por alguna etiqueta que acompañe los datos a procesar. Los modelos de AA no supervisado son preferidos en tareas de procesamiento complejas en las que se requiera organizar grandes conjuntos de datos en grupos o clústeres y son también útiles para identificar patrones de datos no detectados previamente. De forma adicional, pueden identificar características útiles para categorizar la información procesada. Estos modelos reciben datos sin procesar, deben inferir sus propias reglas y estructurar la información en función de similitudes, diferencias y patrones sin instrucciones explícitas de nomenclatura y características de los datos.

Supongamos, por ejemplo, que se tiene un conjunto de datos muy grande sobre el clima. Un modelo de AA no supervisado es capaz de identificar patrones en los datos y agruparlos con base en características de temperatura o datos climáticos similares. Si bien, el modelo no hace una inferencia explícita, los grupos de datos que entrega están separados según los diferentes patrones identificados por el modelo; es decir, será posible reconocer que los patrones climáticos están separados en diferentes tipos de clima, como lluvia, ondas gélidas o nieve.

Este tipo de enfoque de AA también resulta útil en áreas como detección de anomalías, en el que las distintas agrupaciones hechas ayuden a descubrir datos atípicos. Es posible, por ejemplo, explorar datos de transacciones comerciales para descubrir patrones o tendencias para generar recomendaciones personalizadas de ventas o para generar perfiles de clientes agrupando sus rasgos comunes o sus comportamientos de compra. De manera similar, el aprendizaje no supervisado es útil para revelar datos inusuales y ayudar a descubrir eventos o comportamientos que se desvían de los patrones normales, revelando transacciones fraudulentas u otros comportamientos indeseados. Es también útil para agrupar y categorizar textos de artículos, blogs y diversos sitios web para su traducción automática y para la clasificación y agrupación de textos, para después emplearlos en la generación automática de contenido.

Aprendizaje por refuerzo

Inspirado en la psicología del comportamiento, el aprendizaje por refuerzo implica generar un modelo que aprenda a tomar decisiones interactuando con un entorno en el que debe ejecutar alguna tarea (como conducir un vehículo o jugar contra un oponente). El modelo, en función de su desempeño, recibe retroalimentación en forma de recompensas o penalizaciones, lo que le permite refinar y mejorar sus acciones con el tiempo. Esta forma de aprendizaje es similar al proceso de prueba y error observable cuando un niño aprende sobre el entorno que lo rodea.

El aprendizaje por refuerzo se aplica en áreas en las que se deben tomar acciones en un entorno altamente dinámico para obtener el máximo de recompensa acumulada posible. Por ejemplo, un modelo de este tipo se puede utilizar para controlar los semáforos en el tráfico vehicular en zonas urbanas adaptando el control de las luces en función de los volúmenes del tráfico, de horarios y hasta del estado del clima.

Los diferentes métodos de aprendizaje por refuerzo también pueden considerar recompensas no inmediatas, es decir, postergar para obtenerlas. Esta es una estrategia de retraso de recompensa de corto plazo a cambio de maximizar la recompensa final. Estos algoritmos pueden aprender acciones exitosas en entornos en los que las recompensas y penalizaciones estén dadas por un conjunto de reglas predefinidas, o bien, que sean provistas por un supervisor externo. De forma similar, pueden interactuar en entornos donde se aprenda a inferir y estructurar sus propias reglas y decisiones sin la ayuda de un supervisor.

Aprendizaje combinado

Además de las tres categorías anteriores, algunos autores han añadido el aprendizaje semi-supervisado; es decir, cuando el proceso de aprendizaje se realiza combinando una parte supervisada y otra no supervisada. De manera similar, recientemente el aprendizaje profundo (o *deep learning,* por ejemplo, modelos de redes neuronales artificiales con un número muy elevado de parámetros) se ha mencionado como una categoría adicional del AA. Estos modelos han demostrado un rendimiento excepcional en tareas como el reconocimiento de imágenes, el procesamiento del lenguaje natural,

los juegos y la generación de contenido multi-modal (Courville, Goodfellow y Bengio, 2016). Sin embargo, en su aprendizaje emplean, fundamentalmente, combinaciones de los tres principales paradigmas del AA ya mencionados, además de una muy elevada cantidad de datos y de cuantiosos recursos computacionales durante sus etapas de entrenamiento.

Aplicaciones relevantes del aprendizaje automático

1. **Salud.** El AA está revolucionando la atención médica con aplicaciones en el diagnóstico de enfermedades, planes de tratamiento personalizados, descubrimiento de medicamentos y análisis predictivo. Los algoritmos analizan imágenes médicas, datos genómicos y registros electrónicos de salud para mejorar la atención y los resultados para los pacientes (Habehh y Gohel, 2021).
2. **Finanzas.** En el sector financiero, el AA se utiliza para la detección de fraudes, evaluación de riesgos, operaciones de negociación algorítmica y gestión de relaciones con los clientes. Los modelos predictivos analizan las tendencias del mercado, optimizan las carteras de inversión y mejoran los procesos de toma de decisiones (John *et al.,* 2021).
3. **Vehículos autónomos.** Los algoritmos de AA desempeñan un papel fundamental al permitir que los vehículos autónomos perciban su entorno, tomen decisiones en tiempo real y naveguen de manera segura. La visión por computadora, la fusión de sensores y el aprendizaje por reforzamiento contribuyen al desarrollo de tecnologías de conducción autónoma (Yang *et al.,* 2021).
4. **Procesamiento del lenguaje natural (PLN):** el PLN facilita que las máquinas comprendan, interpreten y generen lenguaje humano. Las aplicaciones incluyen traducción de idiomas, análisis de sentimientos, *chat bots* y reconocimiento de voz, mejorando la comunicación entre humanos y máquinas (Bharadiya, 2023).
5. **Manufactura e industria 4.0:** el AA optimiza los procesos de fabricación al predecir fallas de equipos, automatizar el control de calidad y optimizar los horarios de producción. La industria 4.0 utiliza el AA para crear fábricas inteligentes y conectadas con una eficiencia mejorada y menos tiempo de inactividad (Rai *et al.,* 2021).

6. **Comercio electrónico y sistemas de recomendación:** el AA impulsa algoritmos de recomendación que analizan el comportamiento y las preferencias del usuario para sugerir productos, películas o contenido. Las recomendaciones personalizadas mejoran la experiencia del usuario y aumentan la participación en plataformas de comercio electrónico y servicios de transmisión (Lucas *et al.,* 2021).
7. **Monitoreo ambiental:** el AA contribuye al monitoreo ambiental mediante el análisis de imágenes satelitales, datos de sensores y modelos climáticos. Las aplicaciones incluyen la detección de deforestación, el monitoreo de la vida silvestre y la predicción de desastres naturales (*Nature Sustainability,* 2018).

Desafíos y futuras direcciones

1. **Consideraciones éticas:** a medida que las aplicaciones de AA se vuelven más ubicuas, las consideraciones éticas sobre sesgos, equidad, transparencia y privacidad se vuelven cruciales. Abordar estos problemas es esencial para garantizar el desarrollo y la implementación responsables de la IA.
2. **Interpretabilidad y explicabilidad:** a pesar de sus capacidades excepcionales, algunos modelos de AA, en especial los modelos de aprendizaje profundo carecen de interpretabilidad. Construir modelos que sean transparentes y explicables es esencial, especialmente en aplicaciones críticas como la atención médica y las finanzas.
3. **Calidad y cantidad de datos:** los modelos de AA dependen, en gran medida, de la calidad y cantidad de datos. Asegurar conjuntos de datos diversos, representativos y limpios es un desafío continuo, en especial, en dominios donde obtener datos etiquetados es intensivo en recursos.
4. **Generalización y sobreajuste:** lograr una generalización robusta, donde un modelo funcione bien con datos no vistos, es un desafío persistente. Equilibrar la complejidad del modelo para evitar el *overfitting* o sobreajuste (por ejemplo, especializarse únicamente en los datos de entrenamiento) o el *underfitting* o subajuste sigue siendo una consideración crucial.

5. **Seguridad y ataques adversarios:** los modelos de AA son susceptibles a ataques, en los que manipulaciones sutiles a los datos de entrada pueden engañar al modelo. Desarrollar modelos robustos que puedan resistir manipulaciones intencionales es un área de interés para la investigación en IA.

Tendencias actuales en aprendizaje automático

1. **Predominio del aprendizaje profundo:** el aprendizaje profundo ha experimentado un éxito sin precedentes en años recientes. Las redes neuronales profundas han demostrado un desempeño notable en el reconocimiento de imágenes, el procesamiento del lenguaje natural y tareas complejas de reconocimiento de patrones.
2. **Transferencia de aprendizaje:** la transferencia de aprendizaje ha ganado importancia como técnica para aprovechar el conocimiento adquirido para mejorar el desempeño en una tarea diferente, aunque relacionada. Esta tendencia permite que los modelos generalicen mejor en una variedad de dominios, en particular, en situaciones con datos etiquetados limitados.
3. **IA explicable:** a medida que los sistemas de aprendizaje automático se integran cada vez más en los procesos de toma de decisiones, ha crecido la demanda de interpretabilidad y transparencia. La IA explicable tiene como objetivo hacer que los modelos de AA sean más comprensibles, responsables y confiables. La complejidad de esta tarea, sin embargo, es considerablemente elevada.
4. **Avances en el aprendizaje por refuerzo:** el aprendizaje por refuerzo ha experimentado avances significativos. Las aplicaciones van desde algoritmos de juegos hasta sistemas de control robótico, ofreciendo avances potenciales en sistemas autónomos y su impacto y utilidad continúan en aumento.
5. **AutoML y optimización de híper-parámetros:** la automatización en la selección de modelos de aprendizaje automático y el ajuste de híper-parámetros, conocida como AutoML, es una tendencia de interés elevado. Esta permite a los no expertos aprovechar el AA sin necesidad de conocimientos técnicos profundos.

6. **Edge Computing para AA:** la implementación de modelos de aprendizaje automático directamente en varios dispositivos (como IoT o teléfonos móviles) es cada vez más frecuente. Esta tendencia aborda desafíos relacionadas con la privacidad, la latencia y el ancho de banda, aunque también permite el procesamiento local sin depender de servidores centralizados.
7. **Redes generativas adversarias:** estos modelos se han convertido en una herramienta poderosa para generar datos sintéticos, crear imágenes realistas y mejorar el aumento de datos. Desempeñan un papel crucial en aplicaciones como la síntesis de imágenes, la transferencia de estilos y la creación de contenido en diversas áreas.
8. **Robustez y defensa:** con el uso cada vez mayor del AA en aplicaciones críticas, garantizar la solidez de los modelos contra ataques se ha convertido en una preocupación inminente. Cada vez se trabaja más en el desarrollo de modelos que sean más resistentes a la manipulación intencional.

Conclusión

El aprendizaje automático ha inaugurado una era de posibilidades sin precedentes, revolucionando la forma cómo procesamos información, tomamos decisiones e interactuamos con la tecnología. Sus aplicaciones abarcan diversos campos, desde la salud hasta las finanzas y su impacto continúa creciendo.

El aprendizaje automático está a la vanguardia de la innovación tecnológica e impulsa avances en diversos ámbitos. Las múltiples tendencias actuales muestran la naturaleza dinámica del campo. Sin embargo, a medida que el aprendizaje automático evoluciona y continúa impregnando varios aspectos de nuestras vidas, abordar desafíos relacionados con las implicaciones éticas, la interpretabilidad y la calidad de los datos se vuelve crucial y subraya la necesidad de una investigación y un desarrollo continuos.

Los investigadores, los usuarios y los creadores de políticas aplicables deben colaborar para abordar estos desafíos, garantizando que los sistemas de aprendizaje automático no solo brinden un alto rendimiento, sino que también cumplan con estándares éticos y valores sociales.

El futuro inmediato y de largo plazo del aprendizaje automático promete aún mayores avances y un impacto todavía más profundo. El aprendizaje automático promete estar a la vanguardia de la innovación y la transformación social, siempre que estos desafíos actuales se enfrenten con soluciones innovadoras y con un compromiso de prácticas éticas.

A medida que navegamos por este panorama impulsado por la IA se vuelve imperativo fomentar un desarrollo responsable, asegurando que los beneficios del aprendizaje automático se aprovechen para el bienestar de la humanidad en su conjunto. Al menos ese debería ser su fin último.

Referencias

Bharadiya, J. (2023). A Comprehensive Survey of Deep Learning Techniques Natural Language Processing. *European Journal of Technology.* Vol. 7. Núm. 1, pp. 58-66.

Courville, A., Goodfellow, I., Bengio, Y. (2016). Deep Learning. The MIT Press.

Gaur, J., Goel, A. K., Rose, A. y Bhushan, B. (2019). Emerging Trends in Machine Learning. *2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies* (ICICICT), pp. 881-885. DOI: 10. 1109 / ICICICT46008.2019.8993192

Goodell, J. W., Kumar, S., Lim, W. M., Pattnaik, D. (2021). Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. *Journal of Behavioral and Experimental Finance.* Vol. 32. DOI: https://doi.org/10.1016/j.jbef.2021.100577

Habehh, H., Gohel, S. (2021). Machine Learning in Healthcare. *Current genomics,* 22(4), pp. 291-300, 2021. DOI: https://doi.org/10.2174/1389202922666210705124359

Khanal, S. S., Prasad, P. W. C., Alsadoon, A. y Maag, A. (2020). A systematic review: machine learning based recommendation systems for e-learning. *Education and Information Technologies.* Vol. 25. Núm. 4.

Lucas, M. P., Silveira, E. da, Rosa-Righi, R. da, Stoffel, R., Costa, C. da, Victória-Barbosa, J., Scorsatto, R., Arcot, T. (2021). Machine learning through the lens of e-commerce initiatives: An up-to-date systematic literature review. *Computer Science Review*. Vol. 41. DOI: https://doi.org/10.1016/j.cosrev.2021.100414

Mitchell, T. M. (1997a). *Machine Learning.* McGraw-Hill.

Mitchell, T. M. (1997b). Does Machine Learning Really Work? *Invited paper, AAAI Press, AI Magazine.* Vol. 18. Núm. 3, pp.11-20.

Nature Machine Intelligence. (2023). What's the next word in large language models? [Editorial.] *Nature Machine Intelligence*. Vol. 5, pp. 331-33.

Nature Sustainability. (2018). Machine learning for environmental monitoring. *Nature Sustainability.* Vol. 1. Núm. 10, pp. 583-588.

Ngai, E. W. T., Lui, A. K. H. y Lee, M. C. M. (2022). Impact of artificial intelligence investment on firm value. *Annals of Operations Research*. Vol. 308, pp 373-388, 2022. DOI: 10.1007/s10479-020-03862-8.

Rai, R., Tiwari, M. K., Ivanov, D., Dolgui, A. (2021). Machine Learning in Manufacturing and Industry 4.0 Applications. *International Journal of Production Research*. Vol. 59. Núm. 16, pp. 4773-4778. DOI:10.1080/00207543.2021.1956675

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*. Vol. 3. Núm. 3, pp. 210-229. DOI: 10.1147/rd.33.0210

Schemmer, M., Bartos, A., Spitzer, P., Hemmer, P., Kühl, N., Liebschner, J. y Satzger, G. (2023). Towards Effective Human-AI Decision-Making: The Role of Human Learning in Appropriate Reliance on AI Advice. *International Conference on Information Systems.* DOI: 10.48550/arXiv.2310.02108

Turing, A. M. (1950). *Computing machinery and intelligence.* Magdalene College, pp. 433-60.

Yang, Q., Fu, S., Wang, H. y Fang, H. (2021). Machine-Learning-Enabled Cooperative Perception for Connected Autonomous Vehicles: Challenges and Opportunities. *IEEE Network.* Vol. 35. Núm. 3, pp. 96-101. DOI: 10.1109/MNET.011.2000560

CHATGPT Y APRENDIZAJE

Javier Salas-García
Facultad de Ingeniería, UAEMEX

Rosa María Valdovinos-Rosas
Facultad de Ingeniería, UAEMEX

Resumen

Este capítulo explora el impacto de la IA en la educación, centrándose en la herramienta ChatGPT de OpenAI. Se destaca cómo la IA puede transformar la enseñanza y el aprendizaje al mejorar la productividad, los resultados educativos, la instrucción personalizada y la retroalimentación. Sin embargo, también plantea desafíos éticos y prácticos, como la posibilidad de exacerbar las desigualdades existentes y los sesgos en los algoritmos. ChatGPT es un poderoso modelo de lenguaje que puede generar texto similar al humano, participar en conversaciones naturales y proporcionar apoyo personalizado. Esto lo convierte en una herramienta prometedora para la educación abierta y autodirigida. Utiliza el procesamiento del lenguaje natural (PLN) para entender y responder a la entrada del usuario de manera intuitiva. Sin embargo, su uso también genera preocupaciones sobre la integridad académica, el impacto en los resultados de aprendizaje, la privacidad de datos y los desafíos laborales. Se recomienda que los estudiantes sean vigilantes al utilizar ChatGPT y verifiquen la precisión de las respuestas generadas. También se enfatiza la necesidad de adaptar las prácticas de enseñanza y evaluación para aceptar la realidad de vivir en un mundo con IA libremente disponible. De este modo, las instituciones educativas deben examinar cuidadosamente la efectividad e impactos de integrar herramientas como ChatGPT. En lugar de prohibir su uso, se sugiere entender sus alcances y limitaciones para hacer un uso responsable de ellas en la educación. La IA tiene un gran potencial para revolucionar la enseñanza y el aprendizaje, pero su implementación requiere una cuidadosa consideración de los beneficios y riesgos involucrados.

Introducción a la inteligencia artificial en el aprendizaje

La IA ha introducido nuevas herramientas en el entorno educativo con el potencial de transformar los procesos convencionales de enseñanza y aprendizaje. En los últimos años, los avances tecnológicos en IA han llevado a varios desarrollos significativos en su adopción y uso generalizado. Estos avances han presentado al mundo modelos de generación de contenido poderosos que permiten a los usuarios crear desde productos de medios digitales hasta muestras de escritura, de forma instantánea, a través de consultas de texto simples (Trust *et al.,* 2023) y más recientemente, videos a partir de texto.

La IA tiene una amplia variedad de usos potenciales en la educación, incluyendo la mejora de la productividad, los resultados del aprendizaje, la instrucción personalizada, la retroalimentación instantánea y el compromiso del estudiante. Sin embargo, a pesar de todos sus beneficios, el uso de la IA en la educación también plantea importantes problemas éticos y prácticos. Estos problemas incluyen la posibilidad de que pueda aumentar las desigualdades ya existentes en el sistema educativo y la naturaleza inherente de los algoritmos de IA que son propensos al sesgo (Adiguzel, Kaya y Cansu, 2023).

Se necesita un esfuerzo colaborativo que involucre a educadores, investigadores y responsables de políticas para garantizar el uso ético y responsable de la IA en la educación. Podemos construir un sistema educativo más justo y exitoso que brinde a los niños la enseñanza individualizada, la retroalimentación y el apoyo que necesitan resolviendo los problemas planteados por las tecnologías de IA y utilizando sus ventajas. La investigación ha demostrado que la tecnología basada en IA puede mejorar los resultados del aprendizaje y la motivación. Los programas de tutoría basados en IA pueden mejorar el rendimiento y la motivación de los estudiantes en entornos educativos (Firat, 2023).

Un ejemplo de una herramienta de IA que ha tenido un impacto significativo en la educación es ChatGPT, un poderoso modelo de lenguaje de la empresa OpenAI (OpenAI, 16/10/2024). Con su capacidad para generar texto que imita de cerca el lenguaje humano y su capacidad para mantener múltiples conversaciones en curso, es una herramienta versátil que puede ayudar en la educación abierta al proporcionar apoyo personalizado, dirección y retroalimentación a los aprendices autodidactas, aumentando así la motivación y el compromiso.

La capacidad de dicha IA para entender y responder a la entrada de lenguaje natural es una característica clave. Utiliza el procesamiento del lenguaje natural (PLN) para analizar la entrada del usuario y proporcionar respuestas relevantes, lo que le permite una experiencia conversacional natural e intuitiva. Además, la capacidad que tiene para proporcionar asistencia personalizada e interactiva es importante, ya que puede ajustar sus respuestas en función de la entrada del usuario y ofrecer recomendaciones personalizadas.

A medida que estas tecnologías continúan avanzando es importante que las universidades se adapten y acepten el uso de herramientas de IA de una manera que apoye el aprendizaje de los estudiantes y los prepare para los desafíos de un mundo cada vez más digital (Sullivan, Kelly y Mclaughlan, 2023). Aunque ChatGPT no es la primera tecnología de IA que influye en la educación, parece ser la primera en una ola de nuevas herramientas de IA que requerirán un replanteamiento de la enseñanza y el aprendizaje (Trust *et al.,* 2023).

Dada la rápida evolución y adopción de las tecnologías de IA y el potencial de estas herramientas para influir en casi todos los aspectos de la educación y la sociedad, se recomienda incorporar el examen crítico de la IA de manera más orgánica en todo el currículo del programa de preparación docente para lograr su inclusión en el currículo y prácticas (Trust *et al.,* 2023). Este capítulo se sumerge en el fascinante mundo de la IA en la educación, explorando su definición, potencial, beneficios y desafíos éticos y prácticos.

La segunda sección se adentra en estos desafíos, explorando cómo la IA podría exacerbar las desigualdades y sesgos existentes en el sistema educativo. Se destaca la necesidad de esfuerzos colaborativos entre educadores, investigadores y formuladores de políticas para garantizar que la IA se utilice de manera efectiva y responsable en la educación.

La tercera sección presenta a ChatGPT, se discute si puede ser una herramienta versátil para el apoyo y la retroalimentación personalizados en la educación abierta.

En la cuarta sección, se exploran las características clave de ChatGPT, incluyendo su uso del procesamiento del lenguaje natural (PLN) para entender y responder a la entrada del usuario. Se discute cómo puede proporcionar asistencia relevante y personalizada a los estudiantes, para mejorar su experiencia de aprendizaje.

Finalmente, en la quinta sección, se examina el futuro de la IA en la educación. Se discute cómo las herramientas basadas en IA tienen el potencial de revolucionar la educación y mejorar los resultados y la motivación del aprendizaje. Se destaca la necesidad de más investigación para explorar hasta qué punto la IA puede revolucionar la educación.

A través de un análisis y una discusión reflexiva, se espera que este capítulo arroje luz sobre el potencial y los desafíos de la IA en la educación y cómo pueden ser utilizadas para mejorar la enseñanza y el aprendizaje.

Desafíos éticos y prácticos de la inteligencia artificial en la educación

La IA ha introducido nuevas herramientas en el entorno educativo con el potencial de transformar los procesos convencionales de enseñanza y aprendizaje. Sin embargo, a pesar de todos sus beneficios, el uso de la IA en la educación también plantea importantes problemas éticos y prácticos, que incluyen la posibilidad de que aumenten las desigualdades ya existentes en el sistema educativo y la naturaleza inherente de los algoritmos de IA que son propensos al sesgo.

El auge de la IA ha llevado a un aumento en la creación de contenido generado por IA, como texto, imágenes y videos. Sin embargo, esto también plantea el desafío de poder detectar si el contenido es generado por un humano o por una IA (Uzun, 2023). Por otra parte, el contenido generado con IA tiene el potencial de mejorar la accesibilidad para las personas con discapacidades, por ejemplo, a través de la creación de subtítulos cerrados y transcripciones automáticas (Uzun, 2023)

Además, el lanzamiento de algunas herramientas de IA ha provocado preocupaciones significativas sobre la integridad académica en la educación superior. Sin embargo, algunos comentaristas han señalado que las herramientas de IA generativa pueden mejorar el aprendizaje de los estudiantes, es por ello que los académicos deberían adaptar sus métodos de enseñanza y evaluación para adaptarse a esta tecnología emergente (Sullivan, Kelly y Mclaughlan, 2023).

Un análisis de contenido de los artículos de noticias destacó que la discusión pública y las respuestas universitarias sobre herramientas de IA se han centrado principalmente en las preocupaciones de integridad académica y el diseño innovador

de evaluaciones. La literatura también reveló, hasta ahora, la falta de la voz estudiantil en la conversación y la mención de que hay potencial para que las herramientas de IA mejoren el éxito y la participación de los estudiantes de entornos desfavorecidos (Sullivan, Kelly y Mclaughlan, 2023).

Los académicos y los representantes universitarios deben ser conscientes de los marcos que eligen para discutir cuando interactúan con los medios, ya que la cobertura de noticias puede influir en las normas sociales hacia el comportamiento de trampa de los estudiantes y las percepciones públicas de las universidades (Sullivan, Kelly y Mclaughlan, 2023).

A medida que estas tecnologías continúan avanzando es importante que las universidades se adapten y acepten el uso de herramientas de IA de una manera que apoye el aprendizaje de los estudiantes y los prepare para los desafíos de un mundo cada vez más digital (Sullivan, Kelly y Mclaughlan, 2023). Es aconsejable que los estudiantes sean vigilantes mientras reciben diversas asistencias de herramientas de IA; es beneficioso para ellos verificar, analizar y editar las respuestas generadas por estas aplicaciones para asegurar la precisión de la información (Sok y Heng, 2023).

Es esencial utilizar estas herramientas solo para la generación de ideas y la elaboración de esquemas, no para producir artículos de investigación completos, con el fin de prevenir la mala conducta académica intencional o no intencional (Sok y Heng, 2023).

En esta etapa de la era de la IA, las preocupaciones relacionadas con la propiedad intelectual, con el autor, es decir, si un contenido es humano o generado por IA, podrían ser inútiles e improductivas. Las preocupaciones relacionadas con la integridad académica quizá necesiten concentrarse más en el factor contenido; es decir, si el contenido presentado es válido y confiable (Uzun, 2023). A este respecto, es vital que se distinga la diferencia entre usar herramientas de IA como "fuentes" de información *versus* "asistentes de redacción". En el segundo caso, el autor del documento sigue teniendo las riendas del trabajo y usar IA supone sencillamente un ahorro de tiempo que puede producir textos más fáciles de leer.

Es cierto que existen preocupaciones sobre el mal uso del contenido generado por IA, como la creación de noticias falsas (*fake news*) o propaganda. Sin embargo, debe tenerse en cuenta que estas preocupaciones no se limitan al contenido generado por IA. El contenido generado por humanos también puede ser manipulado y utilizado

para difundir desinformación (Uzun, 2023). Por lo que sigue siendo un problema el uso que se les da a las herramientas, y no las herramientas por sí mismas.

Por lo tanto, antes de cuestionar las tecnologías emergentes, podría ser un enfoque más preciso cuestionar las aplicaciones de las filosofías modernas en cualquier ámbito. Ni las filosofías modernas ni las prácticas que se ejecutan con ellas pueden oponerse a los productos de la tecnología (Uzun, 2023).

Es vital evaluar cómo estas herramientas pueden influir positivamente en la enseñanza y el aprendizaje, al tiempo que se identifican los posibles efectos negativos que surgen en el camino. Al hacerlo, los educadores y otros interesados pueden tomar decisiones informadas sobre el uso de estas tecnologías en entornos educativos y desarrollar estrategias para maximizar los beneficios y minimizar los riesgos (Wardat, 2023).

Una vez que alguien se ha familiarizado con las herramientas de IA y sus capacidades puede decidir si utilizar su potencial o no, sin perder la atención sobre sus posibles efectos negativos. Para lograr esto, las personas pueden ajustar sus procesos arraigados, lo que puede ser difícil debido a una resistencia al cambio (Wardat, 2023).

ChatGPT tiene cinco beneficios principales, como la creación de evaluaciones de aprendizaje, la mejora de la práctica pedagógica, la oferta de tutoría personal virtual, la creación de un esquema y la generación de ideas; sin embargo, también existen riesgos relacionados con problemas de integridad académica, evaluación de aprendizaje injusta, información inexacta y dependencia excesiva de la IA (Sok y Heng, 2023).

Para que las escuelas y universidades mejoren la calidad de la educación, se debe hacer una investigación de acción para examinar más a fondo la efectividad y eficiencia de la integración de esta aplicación inteligente en la educación (Sok y Heng, 2023).

ChatGPT: un modelo de lenguaje potente para la educación

ChatGPT, un *chatbot* de conversación de propósito general desarrollado por la empresa OpenAI, ha demostrado ser una herramienta prometedora en el ámbito educativo. Con su capacidad para generar texto similar al humano y participar en conversaciones naturales y abiertas sobre una amplia gama de temas tiene el

potencial de cambiar la forma en que los estudiantes abordan sus investigaciones y cómo se lleva a cabo la educación (Zhai *et al.,* 2022).

ChatGPT puede ser utilizado para aumentar la participación de los estudiantes en las clases en línea, al proporcionarles actividades interactivas y preguntas que se alinean con el material del curso. Además, puede mejorar la experiencia de aprendizaje al brindarles apoyo personalizado e interactivo. Esto incluye ejercicios y juegos personalizados que se alinean con las necesidades específicas del aprendiz, así como recomendaciones para materiales de aprendizaje y recursos.

También puede actuar como tutor o mentor, mediante retroalimentación y asistencia a lo largo del proceso de aprendizaje, y puede ayudar a los estudiantes autodirigidos a desarrollar sus propios objetivos y estrategias de aprendizaje, y puede ser utilizado como una herramienta para la auto-reflexión y evaluación. Esto puede empoderar a los estudiantes para tomar el control sobre su aprendizaje, su crecimiento y el desarrollo de las habilidades que necesitan.

En su versión GPT-3, alcanzó un millón de usuarios en solo cinco días. Con 175 mil millones de parámetros, GPT-3 puede generar escritura que se asemeja estrechamente al lenguaje humano. Respaldado por el modelo GPT-4, ChatGPT puede participar en múltiples conversaciones en curso, entender y responder a la entrada de lenguaje natural y ofrecer asistencia personalizada e interactiva. Esto hace de ChatGPT una herramienta prometedora para la educación abierta, ya que puede mejorar la independencia y autonomía de los aprendices autodidactas, al tiempo que es práctico y adaptable. Al proporcionar apoyo personalizado, dirección y retroalimentación, tiene el potencial de aumentar la motivación y el compromiso entre los alumnos autodidactas (Firat, 2023).

Entender y responder a la entrada en lenguaje natural es una de sus principales características. Utiliza el procesamiento de lenguaje natural (PLN) para examinar la entrada del usuario y producir respuestas pertinentes. Esto permite a los usuarios conversar de una manera que parece natural e intuitiva, de la misma forma en que lo harían con una persona. La capacidad que tiene para ofrecer a los usuarios asistencia individualizada e interactiva es otro aspecto importante (Firat, 2023).

El desarrollo de herramientas de IA tiene el potencial de alterar por completo cómo los estudiantes abordan sus estudios y el ámbito de la educación. Los programas de tutoría basados en IA pueden mejorar el rendimiento y la motivación de los estudiantes

en los entornos de aprendizaje. Al ofrecer asistencia personalizada e interactiva a los estudiantes, las tecnologías de IA, como los *chatbots*, pueden mejorar la experiencia de aprendizaje y aumentar la participación de los estudiantes en los cursos en línea (Firat, 2023).

El avance reciente en el aprendizaje automático ha resultado en la creación de tecnología sofisticada e innovadora de generación de contenido digital, como la IA generativa. Basándose en las respuestas de los participantes, parece que ChatGPT está siendo reconocido por sus mejoradas capacidades matemáticas y su capacidad para aumentar el éxito educativo (Wardat, 2023). Sin embargo, cabe aclarar que existen otras IA que tienen un mejor desempeño para tareas matemáticas, como Claude de la empresa *Anthropic* (16/10/2024).

Aunque, ciertamente, ChatGPT es un potente modelo de lenguaje, no es infalible. De hecho, ha habido casos en los que los usuarios han intentado deliberadamente engañar al modelo para que genere respuestas incorrectas o sesgadas (Wardat, 2023).

Un estudio llevado a cabo por Wardat (2023) pone énfasis en la necesidad de una nueva filosofía educativa que pueda evolucionar con la introducción de *chatbots* en el aula. También subraya el valor de mejorar las competencias de los profesores y los estudiantes con la tecnología de los *chatbots.*

El impacto de ChatGPT en la educación puede ser enorme, ya que su capacidad puede impulsar cambios en los objetivos de aprendizaje educativo, las actividades de aprendizaje y las prácticas de evaluación (Zhai *et al.,* 2022).

El resultado de distintos estudios sugiere que ChatGPT es una herramienta que permite a los usuarios generar documentos coherentes que deberán ser revisados antes de ser divulgados para los diferentes sectores. La escritura es extremadamente eficiente (de dos a tres horas) e involucra un conocimiento profesional muy limitado por parte del autor (Zhai *et al.,* 2022). El artículo plantea la necesidad de actualizar los objetivos de aprendizaje, para que los usuarios se permitan el uso de las herramientas de IA para apoyarse en la realización de sus tareas, mientras que la educación debería centrarse en mejorar la creatividad y el pensamiento crítico de los estudiantes (Zhai *et al.,* 2022).

ChatGPT también plantea preocupaciones acerca de que los estudiantes exhiban tareas de evaluación. Este documento subraya la necesidad de nuevos formatos de evaluaciones que se centren en la creatividad y el pensamiento crítico que la IA no puede sustituir (Zhai *et al.,* 2022).

Características Clave de ChatGPT

A pesar de las ventajas que ofrece ChatGPT, también existen preocupaciones válidas sobre su uso. Algunas de estas incluyen la integridad académica, su influencia en los resultados de aprendizaje y el desarrollo de habilidades, la limitación de capacidades, las preocupaciones políticas y sociales y los desafíos de la fuerza laboral (Li *et al.,* 2023).

ChatGPT puede ser útil en la educación médica, la investigación y la práctica clínica; sin embargo, las capacidades humanas siguen siendo necesarias (Sallam, 2023). ChatGPT tiene cinco beneficios principales, como la creación de evaluaciones de aprendizaje, la mejora de la práctica pedagógica, la oferta de tutoría personal virtual, la creación de un esquema y la generación de ideas. Sin embargo, existen riesgos relacionados con problemas de integridad académica, evaluación de aprendizaje injusta, información inexacta y dependencia excesiva de la IA (Sok y Heng, 2023).

ChatGPT puede asistir en el diseño de evaluaciones, la producción de ensayos y la traducción de idiomas, pero también permite a los usuarios plantear y responder una variedad de preguntas, resumir textos e interactuar con él como si fueran compañeros. También puede ofrecer un sistema de calificación automática con retroalimentación útil, que es esencial para mejorar los resultados de aprendizaje de los estudiantes. ChatGPT podría ser aprovechado para semiautomatizar la calificación del trabajo de los estudiantes a través de la identificación de las debilidades y fortalezas de la tarea en cuestión (Sok y Heng, 2023).

Para que las escuelas y universidades mejoren la calidad de la educación, se debe hacer una investigación de acción para examinar más a fondo la efectividad y eficiencia de la integración de esta aplicación inteligente en la educación. Mientras reciben diversas asistencias de ChatGPT es aconsejable que los estudiantes se mantengan vigilantes. Quizá es beneficioso para los estudiantes verificar, analizar y editar las respuestas generadas por esta aplicación de IA para asegurar su precisión. Sin embargo, el ChatGPT también presenta desventajas significativas que son señaladas por distintos investigadores. Un estudio (Mhlanga, 2023) encontró que, en una evaluación inicial, ChatGPT no respondió consistentemente con información precisa, en su versión actual, al preguntarle sobre hechos anatómicos (Sok y Heng, 2023).

Los usuarios de X (antes Twitter) por lo general expresan una actitud positiva hacia el uso de ChatGPT, sus preocupaciones convergieron en cinco categorías específicas: integridad académica, impacto en los resultados de aprendizaje y desarrollo de habilidades, limitación de capacidades, preocupaciones políticas y sociales, y desafíos de la fuerza laboral. También se encontró que los usuarios de los campos de la tecnología, la educación y los medios de comunicación, a menudo, estaban implicados en la conversación, mientras que los usuarios individuales de la educación y la tecnología lideraban la discusión de las preocupaciones (Li *et al.,* 2023).

La protección de los datos de los usuarios es una prioridad. Dado que ChatGPT se entrena en enormes volúmenes de datos obtenidos de internet, es esencial garantizar que los datos personales de los estudiantes estén protegidos y no se utilicen de manera inapropiada. Antes de usarlo en el aula, los educadores deben informar a los estudiantes sobre cómo se recopilan, utilizan y mantienen sus datos y obtener su consentimiento. Los estudiantes, por su parte, deben estar al tanto de las medidas de seguridad vigentes para proteger sus datos. En cuanto a la aplicación de ChatGPT en el ámbito de la educación, la protección de la información personal de los usuarios es una preocupación esencial. Es imperativo que siempre se garantice la privacidad y seguridad de los datos de los usuarios (Mhlanga, 2023).

Aplicaciones potenciales de la IA en la educación

La capacidad de herramientas de IA para entender y responder a la entrada de lenguaje natural se siente natural e intuitiva. Su capacidad para proporcionar asistencia personalizada e interactiva es importante, ya que puede ajustar sus respuestas en función de la entrada del usuario y ofrecer recomendaciones personalizadas.

Estas herramientas tienen el potencial de revolucionar la educación y la forma en que los estudiantes abordan sus estudios. La investigación ha demostrado que la tecnología basada en IA puede mejorar los resultados del aprendizaje y la motivación. Los programas de tutoría basados en IA pueden mejorar el rendimiento y la motivación de los estudiantes en entornos educativos.

Las IA se pueden utilizar para aumentar la participación de los estudiantes en las clases en línea, al proporcionarles actividades interactivas y preguntas que se alinean con el material del curso. También puede actuar como tutor o mentor, al brindarles retroalimentación y asistencia a lo largo del proceso de aprendizaje. Esto puede empoderar a los estudiantes para tomar el control de su aprendizaje, su crecimiento y el desarrollo de las habilidades necesarias para el éxito como estudiantes autodirigidos.

ChatGPT puede asistir en el diseño de evaluaciones, la producción de ensayos y la traducción de idiomas, pero también permite a los usuarios plantear y responder una variedad de preguntas, resumir textos e interactuar con él como si fueran compañeros.

Para que las escuelas y universidades mejoren la calidad de la educación se debe hacer una investigación de acción para examinar más a fondo la efectividad y eficiencia de la integración de esta aplicación inteligente en la educación. Mientras se reciben diversas asistencias de herramientas de IA (Sok y Heng, 2023).

Es indudable que las herramientas de IA han llegado para quedarse. El camino que deben seguir las instituciones y docentes a nivel individual no es el de prohibir su uso, sino entender sus alcances y limitaciones, para promover un uso responsable de ellas. Aquí solo se habló en particular de ChatGPT. Sin embargo, cada semana aparecen nuevas. Muchas de ellas son solo aplicaciones que usan modelos ya existentes en el mercado, a traves de una interfaz de programación de aplicaciones (API), es decir, sitios de internet o aplicaciones que toman ciertos datos de una interfaz con el usuario, los envían a modelos de IA para su procesamiento y regresan la salida de dichas API al usuario en dicha interfaz. En otros casos se trata de modelos nuevos que se suman al desarrollo de grandes empresas, como Google, Anthropic, Facebook, Microsoft, etcétera. ¿Cómo navegar en ese mar de posibilidades? ¿Cómo estar al día de los avances mensuales, e incluso semanales, de tantos actores?

Pensemos en qué hacemos al escoger un celular… aunque hay quienes de forma casi obsesiva buscan las características y están al día en las noticias de nuevos lanzamientos de muchas empresas, por lo general, el usuario de término medio lee reseñas y elige uno que se ajuste a sus necesidades. Algo así podría hacerse en este campo. Conocer las características de los principales actores nos permitirá tomar decisiones sobre cuáles tecnologías usar, para qué propósito y en qué medida.

Referencias

Adiguzel, T., Kaya, M. H. y Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology, 15*(3), ep429. DOI: https://doi.org/10.30935/cedtech/13152

Anthropic. (16/10/2024). About Anthropic. *Anthropic.* [En línea.] https://www.anthropic.com/

Firat, M. (2023). *How ChatGPT Can Transform Autodidactic Experiences and Open Education*. DOI: https://doi.org/10.31219/osf.io/9ge8m

Li, L., Ma, Z., Fan, L., Lee, S., Yu, H. y Hemphill, L. (2023). *ChatGPT in education: A discourse analysis of worries and concerns on social media.* DOI: arXiv:2305.02201. https://doi.org/10.48550/arXiv.2305.02201

Mhlanga, D. (2023). Open AI in Education, the Responsible and Ethical Use of ChatGPT Towards Lifelong Learning. [Artículo académico 4354422.] DOI: https://doi.org/10.2139/ssrn.4354422

OpenAI. (16/10/2024). About OpenAI. *OpenAI.* [En línea.] https://openai.com/about/

Sallam, M. (2023). ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare, 11*(6). Article 6. DOI: https://doi.org/10.3390/healthcare11060887

Sok, S. y Heng, K. (2023). *ChatGPT for Education and Research: A Review of Benefits and Risks.* [Artículo académico 4378735.] DOI: https://doi.org/10.2139/ssrn.4378735

Sullivan, M., Kelly, A. y Mclaughlan, P. (2023). *ChatGPT in higher education: Considerations for academic integrity and student learning. Journal of Applied Learning Teaching*. DOI: https://doi.org/10.37074/jalt.2023.6.1.17

Trust, T., Krutka, D. G., Carpenter, J. P., Kimmons, R. y Miller, K. (2023). Emerging technologies: The role of ChatGPT in K-12 classrooms. *Journal of Technology and Teacher Education, 31*(2), pp. 233-259. DOI: https://www.learntechlib.org/p/231031

Uzun, E. (2023). ChatGPT in education: A paradigm shift or a passing fad?. *TechTrends, 67*(3), pp. 320-332. DOI: https://doi.org/10.1007/s11528-023-00598-y

Wardat, S. (2023). *ChatGPT: Possibilities, challenges and ethical considerations in education.* [Artículo académico 4375453.] DOI: https://papers.ssrn.com/abstract=4375453

Zhai, C., Fang, H., Lu, Z., Zhang, Y. y Gao, X. (2022). Exploring ChatGPT for conversational search. *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pp. 529-537. DOI: https://doi.org/10.1145/3469327.3478795

HACIA UN FUNDAMENTO MATEMÁTICO DE LA INTELIGENCIA ARTIFICIAL

José Raymundo Marcial Romero
Facultad de Ingeniería, UAEMEX

Introducción

En la actualidad, la IA está experimentando avances significativos, como en el desarrollo de sistemas de conducción autónoma y reconocimiento de voz, así como en su aplicación en campos científicos, como el diagnóstico médico y el análisis de la dinámica molecular. Asimismo, la investigación en IA,en especial, en sus bases teóricas, está progresando a un ritmo sin precedentes. Es concebible que, en el futuro, estas tecnologías transformen radicalmente nuestra manera de vivir en múltiples aspectos.

La IA no es un campo nuevo. En 1943, McCulloch y Pitts comenzaron a desarrollar métodos algorítmicos para imitar el funcionamiento del cerebro humano mediante neuronas artificiales interconectadas en varias capas para formar redes neuronales artificiales (Chakraverty, Sahoo y Mahato, 2019). En aquel entonces, tenían la visión de implementar la IA. Sin embargo, la comunidad no comprendió completamente el potencial de las redes neuronales, lo que llevó al fracaso de esta primera ola de IA, que eventualmente desapareció.

Hacia 1980, el aprendizaje automático volvió a ganar popularidad, y se pueden mencionar varios hitos importantes en ese periodo (Cybenko, 1989; Hornik, Stinchcombe y White, 1989).

El verdadero avance que marcó el comienzo de una nueva era en la IA ocurrió alrededor de 2010, con la amplia aplicación de las redes neuronales profundas (Berner *et al.,* 2022). En la actualidad, este modelo podría considerarse como el pilar fundamental de la IA, en este capítulo se abordará principalmente este enfoque. La estructura de las redes neuronales profundas es en esencia la misma que introdujeron McCulloch y Pitts, que consiste en numerosas capas consecutivas de neuronas artificiales (Cybenko, 1989; Bölcskei *et al.,* 2019). Hoy en día, también se han superado dos obstáculos principales de años anteriores: gracias al significativo avance

en la capacidad de procesamiento computacional, ahora es posible entrenar redes neuronales con cientos de capas, en lo que se refiere a las redes neuronales profundas, y se está inmerso en la era de los datos, lo que significa que grandes volúmenes de datos de entrenamiento están fácilmente disponibles (Donoho, 2001).

El surgimiento de la IA también ha tenido un impacto considerable en diversos ámbitos de las matemáticas. Uno de los primeros campos en adoptar estos enfoques innovadores fue el de los problemas inversos, en especial, en la ciencia de la imagen, donde se han empleado para abordar problemas como la eliminación de ruido, la superresolución o la tomografía computarizada (de ángulo limitado), entre otros. Se puede notar que, debido a la falta de un modelo matemático preciso de lo que constituye una imagen, este campo es en particular propicio para los métodos de aprendizaje. Como resultado, en el transcurso de unos años se ha observado un cambio de paradigma, y los nuevos enfoques de resolución suelen basarse, al menos en cierta medida, en métodos de IA (E y Yu, 2018; Geist *et al.,* 2021).

El campo de las ecuaciones diferenciales parciales tardó más en adoptar estas nuevas técnicas, principalmente, porque no estaba claro cuál sería la ventaja de utilizar métodos de IA en este campo. De hecho, parecía no haber necesidad de recurrir a métodos de aprendizaje, dado que una ecuación diferencial parcial es un modelo matemático riguroso. Sin embargo, recientemente la observación de que las redes neuronales profundas pueden manejar la dimensionalidad en entornos de alta dimensión ha llevado a un cambio de paradigma, también en este campo. Como resultado, la investigación en la intersección del análisis numérico de ecuaciones diferenciales parciales y la IA ha experimentado un aumento considerable desde el año 2017 (Han, Jetzen y Weinan, 2018; Jin *et al.,* 2017).

Desafíos en la inteligencia artificial

A pesar de la aparente promesa de todos estos avances es esencial hacer una advertencia. Además de las limitaciones prácticas que aún no se han explorado completamente, como las asociadas con los métodos como redes neuronales profundas, que todavía se consideran "expertos en todas las áreas", es aún más preocupante la falta de una base teórica sólida (Kolek *et al.,* 2022). Esta preocupación fue destacada de manera

significativa durante la principal conferencia sobre IA y aprendizaje automático, NIPS (ahora llamada NeurIPS), en 2017, cuando Ali Rahimi de Google recibió el premio Test of Time y afirmó durante su charla plenaria que "el aprendizaje automático se ha convertido en una forma de alquimia" (Rahimi, 2007). Esta declaración desencadenó un intenso debate sobre la existencia y necesidad de una base teórica sólida. Desde una perspectiva matemática, es evidente que una comprensión matemática fundamental de la IA es inevitable y se debe reconocer que, en la actualidad, su desarrollo se encuentra, en el mejor de los casos, en una etapa preliminar.

La ausencia de fundamentos matemáticos sólidos, sobre todo en el caso de las redes neuronales profundas, impulsa a una búsqueda de una arquitectura de red adecuada, un proceso de entrenamiento muy delicado que se basa en ensayo y error, así como la ausencia de límites de error definidos para evaluar el rendimiento de la red neuronal entrenada. Además, es crucial destacar que estos enfoques a veces fallan de manera inesperada, cuando una pequeña alteración en los datos de entrada provoca un cambio drástico en la salida, lo que conduce a decisiones radicalmente diferentes y, en muchas ocasiones, incorrectas (Kutyniok *et al.*, 2022). Estos ejemplos contradictorios representan un problema bien conocido, en particular, grave en aplicaciones sensibles, como cuando incluso pequeñas modificaciones en las señales de tráfico, como la colocación de *stickers,* pueden llevar a los vehículos autónomos a tomar decisiones, de manera repentina, completamente equivocadas (Lundberg y Lee, 2017). Es evidente que estos problemas de robustez solo pueden abordarse mediante un análisis matemático exhaustivo.

Una demanda por mayores conocimientos matemáticos

En la actualidad, numerosos matemáticos están incursionando en este campo, aportando sus propios conocimientos (Lunz, Öktem y Schönlieb, 2018; Raghu *et al.,* 2017). Es evidente que prácticamente todas las áreas de las matemáticas son necesarias para abordar los diversos desafíos, tanto difíciles como apasionantes, en el campo de la inteligencia artificial.

Se pueden distinguir dos enfoques de investigación diferentes en la intersección de las matemáticas y la inteligencia artificial:

Los fundamentos matemáticos de la IA buscan alcanzar una comprensión matemática profunda. Su objetivo principal es superar los desafíos actuales, como la falta de robustez, y establecer todo el proceso de aprendizaje en bases teóricas sólidas (Raissi, Perdikaris y Karniadakis, 2019).

La IA aplicada a problemas matemáticos se centra en resolver desafíos matemáticos específicos, como problemas inversos y ecuaciones diferenciales parciales. Su propósito es utilizar enfoques de IA para crear solucionadores más eficaces (Reisenhofer, Kiefer y King, 2015).

El panorama matemático de la inteligencia artificial

Ya que el enfoque que discute este capítulo se basa en redes neuronales profundas, se presenta su definición. Asimismo, se presenta la configuración típica de su aplicación y el proceso de entrenamiento, así como las principales áreas matemáticas de interés en la actualidad.

Redes neuronales profundas

Como se mencionó antes, los elementos fundamentales son las neuronas artificiales. Para comprender su definición es útil recordar la estructura y función de una neurona en el cerebro humano. Una neurona típica consta de dendritas, que reciben señales y las transmiten al soma, en donde se acumulan y se procesan estas señales entrantes. Luego, en el soma se toma la decisión de si la neurona disparará señales a otras neuronas y con qué intensidad lo hará.

Esto constituye la base para una definición matemática de neurona artificial.

Definición. Una neurona artificial con pesos $w_1, \dots, w_n \in \mathbb{R}$, bias $b \in \mathbb{R}$ y función de activación $\rho: \mathbb{R} \longrightarrow \mathbb{R}$ se define como la función $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ dada por

$$f(x_1, \dots, x_n) = \rho\left(\sum_{i=1}^{n} x_i w_i - b\right) = \rho(\langle x, w \rangle - b)$$

Donde $w=(w_1,\ldots,w_n)$ y $x=(x_1,\ldots,x_n)$

En la actualidad, existe una gama de funciones de activación, las más conocidas son las siguientes:

1. Función escalón $\rho(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}$
2. Función sigmoid $\rho(x) = \dfrac{1}{1+e^{-x}}$
3. Unidad lineal rectificable (ReLU) $\rho(x) = \max\{0, x\}$

De acuerdo con la revisión literaria realizada, la función de activación más utilizada es ReLU debido a su estructura lineal simple por partes, que es ventajosa en el proceso de entrenamiento y aún permite un rendimiento superior.

De manera similar a la estructura de un cerebro humano, estas neuronas artificiales ahora se concatenan y organizan en capas, lo que crea una red neuronal. Debido a la estructura particular de las neuronas artificiales, dicha red neuronal consta de composiciones de mapas lineales afines y funciones de activación. Desde un punto de vista matemático, esto presenta la dificultad de que diferentes disposiciones conducen a la misma función. Por lo tanto, a veces se hace una distinción entre la arquitectura de una red neuronal y la función de activación correspondiente (véase Berner *et al.,* 2022).

Definición. Sea $d \in \mathbb{N}$ la dimensión de la capa de entrada, L el número de capas, $N_0 := d, N_l, l = 1, \ldots L$ las dimensiones de las capas ocultas y capa de salida, respectivamente, $\rho: \mathbb{R} \to \mathbb{R}$ una función de activación (no lineal), y para $l = 1, \ldots, L$ sea T_l las funciones afines lineales

$$T_l: \mathbb{R}^{N_{l-1}} \to \mathbb{R}^{N_l}, \quad T_l(x) = W^{(l)}x + b^{(l)}$$

Con $W^l \in \mathbb{R}^{N_l * N_{l-1}}$ como los pesos de las matrices y $b^{(l)} \in \mathbb{R}^{N_l}$ el vector de la l-ésima capa. Por lo tanto $\Phi: \mathbb{R}^d \to \mathbb{R}^{N_L}$, dado por

$$\Phi(x) = T_L\rho(T_{L-1}\rho(\dots\rho(T_1))), \qquad x \in \mathbb{R}^d$$

Se llama red neuronal de profundidad *L*. Los pesos y sesgos son los parámetros libres que se aprenden durante el proceso de formación.

Principales líneas de investigación

Se pueden identificar dos líneas de investigación conceptualmente diferentes: la primera centrada en desarrollar fundamentos matemáticos de la IA y, la segunda, orientada a utilizar metodologías de la IA para resolver problemas matemáticos. Ambas, ya han conducido, hasta cierto punto, a un cambio de paradigma en algunas áreas de investigación matemática, sobre todo en el área del análisis numérico (Ribeiro, Singh y Guestrin, 2016; Robbins y Monro, 1952; Romano, Elad y Milanfar, 2017).

Fundamentos matemáticos de la inteligencia artificial

Existen al menos tres direcciones de investigación que están relacionadas con los tres tipos de errores que es necesario controlar para estimar el error general de todo el proceso de entrenamiento.

Expresividad. Esta dirección tiene como objetivo obtener una comprensión general de la afectación de los aspectos de una arquitectura de red neuronal sobre el rendimiento de las redes neuronales profundas y en qué medida lo hacen. Los métodos típicos para abordar este problema provienen del análisis armónico aplicado y la teoría de aproximación (Rumelhart, Hinton y Williams, 1986).
Aprendizaje/optimización. El objetivo principal de esta dirección es el análisis de algoritmos de entrenamiento, como el descenso de gradiente estocástico, en particular, preguntando por qué normalmente converge a mínimos locales adecuados a pesar de que el problema en sí es altamente no convexo.

Las metodologías clave para resolver estos problemas provienen de las áreas de geometría algebraica/diferencial, control óptimo y optimización (Smilkov *et al.,* 2017).

Generalización. Esta dirección tiene como objetivo obtener una comprensión del error fuera de muestra. La teoría del aprendizaje, la teoría de la probabilidad y la estadística proporcionan, de forma predominante, los métodos necesarios para este hilo de investigación (Ulyanov, Vedaldi y Lempitsky, 2018).

Recientemente ha surgido una nueva dirección de investigación interesante y relevante: la explicabilidad. En la actualidad, desde el punto de vista de los fundamentos matemáticos, sigue siendo un campo abierto.

Explicabilidad. Esta dirección considera redes neuronales profundas, que ya están entrenadas, pero no se dispone de conocimiento sobre el entrenamiento; una situación que uno encuentra numerosas veces en la práctica. El objetivo entonces es obtener una comprensión de cómo una determinada red neuronal profunda entrenada toma decisiones, es decir, qué características de los datos de entrada son cruciales para una decisión. La gama de enfoques necesarios es bastante amplia e incluye áreas como la teoría de la información o la cuantificación de la incertidumbre (Wäldchen *et al.,* 2021).

Inteligencia artificial para problemas matemáticos

Los métodos de IA también han resultado extremadamente eficaces para la resolución de problemas matemáticos. De hecho, el área de los problemas inversos, en particular en las ciencias de la imagen, ya ha experimentado un profundo cambio de paradigma. Y el área del análisis numérico de ecuaciones diferenciales parciales parece seguir el mismo camino, al menos cuando se consideran muy altas dimensiones.

Problemas inversos. La investigación en esta dirección tiene como objetivo mejorar los enfoques clásicos basados en modelos para resolver problemas inversos mediante la explotación de métodos de IA. Para no descuidar el conocimiento de dominios, como la física del problema, los enfoques actuales apuntan a obtener lo mejor de ambos mundos, esto significa combinar de manera óptima enfoques basados en modelos y datos. Esta dirección de investigación requiere una variedad de técnicas, principalmente de áreas como la ciencia de la imagen, los problemas inversos y el análisis microlocal, por nombrar algunas (Lucas *et al.,* 2018; Adler y Öktem, 2017).

Ecuaciones diferenciales parciales. Al igual que en el área de problemas inversos, aquí el objetivo es mejorar los solucionadores clásicos de ecuaciones diferenciales

parciales utilizando ideas de IA. Se da especial atención a los problemas de altas dimensiones con el propósito de intentar vencer la dimensionalidad. Obviamente, esta dirección requiere métodos de áreas como las matemáticas numéricas y las ecuaciones diferenciales parciales.

En este capítulo se han presentado las redes neuronales profundas como uno de los enfoques para la IA; sin embargo, a pesar de los excelentes resultados prácticos obtenidos falta mucho por encontrar el fundamento formal (matemático) para abordarlo.

Referencias

Adler, J. y Öktem, O. (2017). Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Probl. 33,* p. 124007.

Berner, J., Grohs, P., Kutyniok, G. y Petersen, P. (2022). The Modern Mathematics of Deep Learning. *Mathematical Aspects of Deep Learning*. Cambridge University Press.

Bölcskei, H., Grohs, P., Kutyniok, G. y Petersen, P. (2019). Optimal Approximation with Sparsely Connected Deep Neural Networks. *SIAM J. Math. Data Sci. 1,* pp. 8-45.

Chakraverty, S., Sahoo, D. M., Mahato, N. R. (2019). Neural Network Model. *Concepts of Soft Computing*. Springer.

Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. Math. *Control Signal 2,* pp. 303-314.

Donoho, D. (2001). Sparse components of images and optimal atomic decompositions. *Constr. Approx. 17,* pp. 353-382.

E, W. y Yu, B. (2018). The deep ritz method: a deep learning-based numerical algorithm for solving variational problems. *Commun. Math. Stat. 6,* pp. 1-12.

Geist, M., Petersen, P., Raslan, M., Schneider, R. y Kutyniok, G. (2021). Numerical Solution of the Parametric Diffusion Equation by Deep Neural Networks. *J. Sci. Comput. 88.* Article núm. 22.

Han, J., Jentzen, A. y Weinan, E. (2018). Solving high-dimensional partial differential equations using deep learning. *Proc. Natl. Acad. Sci. USA 115,* pp. 8505-8510.

Hornik, K., Stinchcombe, M. y White, H. (1989). Multilayer feed forward networks are universal approximators. *Neural Netw. 2,* pp. 359-366.

Jin, K. H., McCann, M. T., Froustey, E. y Unser, M. (2017). Deep convolutional neural network for inverse problems in imaging. *IEEE Trans. Image Process. 26,* pp. 4509-4522.

Kolek, S., Nguyen, D. A., Levie, R., Bruna, J. y Kutyniok, G. (2022). A rate-distortion framework for explaining black-box model decisions. *Springer LNAI.* Vol. 13200, xxAI. Beyond Explainable AI.

Kutyniok, G., Petersen, P., Raslan, M. y Schneider, R. (2022). A Theoretical Analysis of DeepNeural Networks and Parametric PDEs. *Constr. Approx. 55,* pp. 73-125.

Lucas, A., Iliadis, M., Molina, R. y Katsaggelos, A. K. (2018). Using Deep Neural Networks for Inverse Problems in Imaging: Beyond Analytical Methods. *IEEE Signal Processing Magazine*. Vol. 35. Núm. 1, pp. 20-36.

Lundberg, S. M. y Lee, S.-I. (2017). A unified approach to interpreting model predictions. *NeurIPS,* pp. 4768-4777.

Lunz, S., Öktem, O. y Schönlieb, C.-B. (2018). Adversarial regularizers in inverse problems. *NeurIPS,* pp. 8507-8516.

Raghu, M., Poole, B., Kleinberg, J., Ganguli, S. y Sohl-Dickstein, J. (2017). On the expressive power of deep neural networks. *ICML,* pp. 2847-2854.

Rahimi, B. R. A. (2007). *Random Features for Large-Scale Kernel Machines.* NIPS.

Raissi, M., Perdikaris, P. y Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning frame- work for solving forward and inverse problems involving nonlinear pdes. *J. Comput. Phys. 378,* pp. 686-707.

Reisenhofer, R., Kiefer, J. y King, E. J. (2015). Shearlet-based detection of flame fronts. *Exp. Fluids 57,* 11.

Ribeiro, M. T., Singh, S. y Guestrin, C. (2016). Why should I trust you?: Explaining the predictions of any classifier. *ACM SIGKDD,* pp. 1135-1144.

Robbins, H. y Monro, S. (1952). A stochastic approximation method. *Ann. Math. Statist. 22,* pp. 400-407.

Romano, Y., Elad, M. y Milanfar, P. (2017). The little engine that could: Regularization by denoising. *SIAM J. Imaging Sci. 10,* pp. 1804-1844.

Rumelhart, D. E., Hinton, G. E. y Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature 323,* pp. 533-536.

Smilkov, D., Thorat, N., Kim, B. y Vi'egas, F. y Wattenberg, M. (2017). Smoothgrad: removing noise by adding noise. *ICML 2017.* Workshop on Visualization for Deep Learning.

Ulyanov, D., Vedaldi, A. y Lempitsky, V. (2018). Deep image prior. *CVPR 2018,* pp. 9446-9454.

Wäldchen, S., Macdonald, J., Hauch, S. y Kutyniok, G. (2021). The computational complexity of under-standing network decisions. *J. Artif. Intell. Res. 70,* pp. 351-387.

Marco Antonio Ramos Corchado es doctor en Inteligencia Artificial Distribuida por la Université de Toulouse 1, Francia. En la actualidad es profesor investigador en la Universidad Autónoma del Estado de México (UAEMEX), adscrito a la Facultad de Ingeniería. Forma parte del Sistema Nacional de Investigadores (SNI), Nivel I. Sus intereses científicos incluyen inteligencia artificial, neurocomputación, realidad virtual y algoritmia.

Vianney Muñoz Jiménez es doctora en Redes y Tecnología de la Información por la Université Paris 13, Francia. Actualmente es profesora investigadora en la UAEMEX, adscrita a la Facultad de Ingeniería. Es integrante del SNI, Nivel I. Sus intereses científicos incluyen procesamiento de imágenes, visión computacional, inteligencia artificial, neurocomputación y realidad virtual.

Inteligencia artificial: teoría y aplicaciones es el resultado del trabajo colaborativo del cuerpo académico Sistemas Computacionales de la Facultad de Ingeniería, de la UAEMEX, su objetivo principal es dar a conocer al público general las investigaciones que persigue cada uno de sus integrantes con el uso de la inteligencia artificial, herramienta que hoy está siendo utilizada en la mayoría de los sistemas aplicativos a los que puede acceder el usuario de forma cotidiana.

En esta obra se tratan temas relacionados con el uso de algoritmos de la inteligencia artificial y algunas de sus bases y aplicaciones, por lo que el lector podrá ampliar su panorama sobre la importancia de la inteligencia artificial, sus retos y sus limitaciones, la manera en que se formaliza y las tendencias futuras que podrán ser implementadas en beneficio de la sociedad. En el ámbito académico la obra, en cada uno de sus capítulos, ofrece pautas y formalismos para poder aplicar la inteligencia artificial de manera responsable, en el futuro, en la resolución de problemáticas particulares o generales en beneficio de los diferentes sectores del país.

SDC