

Inteligencia Computacional



“Actualmente es muy frecuente **vincular la Inteligencia Artificial solo al análisis (¿inteligente?) de datos**, y éste a las técnicas de 'aprendizaje automático' o 'machine learning', en su mayoría provenientes de la estadística, más o menos básica. Pero la Inteligencia Artificial es mucho más que eso, más aún, algo muy diferente: **dotar a sistemas computacionales con la capacidad de simular comportamientos inteligentes humanos, tanto en raciocinio como en actuación.**

Por otra parte, asistimos al fenómeno del '**endiosamiento**' de los datos, cuando sabemos que son representaciones muy **incompletas, mayoritariamente numéricas, con ruido, imprecisión, incertidumbre y cada vez más, deliberadamente sesgadas o falsas, habitualmente no con las mejores intenciones.**”

José Ángel Olivas Varela

La Inteligencia Artificial (IA/AI –siglas en inglés–) puede ser definida como la disciplina del ámbito de la computación y los sistemas de información con el objetivo de simular comportamientos humanos que pueden ser considerados como inteligentes, tanto en cuanto a actuación como a razonamiento, en el sentido de ser capaz de generar inferencias.

¿Qué es la Inteligencia Computacional?

El campo de interés de la [Sociedad de Inteligencia Computacional](#) (CIS en inglés) es la teoría, el diseño, la aplicación y el desarrollo de paradigmas computacionales motivados biológica y lingüísticamente. La Sociedad de Inteligencia Computacional forma parte del [Instituto de Ingenieros Eléctricos y Electrónicos \(IEEE\)](#) es una asociación mundial de ingenieros dedicada a la normalización y el desarrollo en áreas técnicas.

Tradicionalmente, los tres pilares principales de la CI han sido las **redes neuronales**, los **sistemas difusos** y la **computación evolutiva**. Sin embargo, con el tiempo han evolucionado muchos paradigmas informáticos inspirados en la naturaleza. Por lo tanto, **CI es un campo en evolución** y, en la actualidad, además de los tres componentes principales, abarca paradigmas informáticos como **inteligencia ambiental, vida artificial, aprendizaje cultural, redes endocrinas artificiales, razonamiento social y redes de hormonas**

artificiales. CI juega un papel importante en el desarrollo de sistemas inteligentes exitosos, incluidos juegos y sistemas de desarrollo cognitivo.

La expresión inteligencia computacional (IC) generalmente se refiere a la capacidad de una computadora para aprender una tarea específica a partir de datos u observaciones experimentales. Aunque comúnmente se considera un sinónimo de computación blanda, todavía no existe una definición comúnmente aceptada de inteligencia computacional.

La noción de Inteligencia Computacional fue utilizada por primera vez por el IEEE Neural Networks Council en 1990. La IEEE Computational Intelligence Society se crea en 2003 al incluir nuevas áreas de interés como los sistemas difusos y la computación evolutiva, que se relacionaron con la Inteligencia Computacional en 2011.

La primera definición clara de Inteligencia Computacional fue presentada en 1994: "un sistema se llama computacionalmente inteligente si maneja datos de bajo nivel como datos numéricos, tiene un componente de reconocimiento de patrones y no usa conocimiento en el sentido de la IA y, además, cuando comienza a exhibir adaptabilidad computacional, tolerancia a fallas, velocidad que se acerca a la respuesta humana y tasas de error que se aproximan al rendimiento humano".

En los últimos años ha habido una explosión de investigación sobre **Deep Learning**, en particular redes neuronales convolucionales profundas. Hoy en día, el aprendizaje profundo se ha convertido en el método central de la inteligencia artificial. De hecho, algunos de los sistemas de IA más exitosos se basan en CI.

En general, la inteligencia computacional es un conjunto de metodologías y enfoques computacionales inspirados en la naturaleza para abordar problemas complejos del mundo real para los cuales los modelos matemáticos tradicionales pueden ser inútiles por varias razones: los procesos pueden ser demasiado complejos para el razonamiento matemático, pueden contener incertidumbre o imprecisión, o el proceso puede ser simplemente de naturaleza estocástica.

Los métodos utilizados en CI se acercan a la forma de razonar del ser humano, es decir, utiliza conocimientos inexactos e incompletos, y es capaz de producir acciones de control de forma adaptativa.

La Inteligencia Computacional es, pues, una forma de actuar como seres humanos. De hecho, la característica de "inteligencia" suele atribuirse a los humanos. Más recientemente, muchos productos y artículos también afirman ser "inteligentes", atributo que está directamente relacionado con el razonamiento y la toma de decisiones.

Si bien la Inteligencia Artificial y la Inteligencia Computacional buscan un objetivo similar a largo plazo: alcanzar la inteligencia general, que es la

inteligencia de una máquina que podría realizar cualquier tarea intelectual que un ser humano pueda realizar; hay una clara diferencia entre ellos.

Existen dos tipos de inteligencia artificial: la artificial basada en técnicas de hard computing y la computacional basada en métodos de soft computing, que permiten la adaptación a multitud de situaciones.

Las técnicas de hard computing funcionan siguiendo una lógica binaria basada en solo dos valores (los booleanos verdadero o falso, 0 o 1) en los que se basan las computadoras modernas. Un problema con esta lógica es que nuestro lenguaje natural no siempre se puede traducir fácilmente a términos absolutos de 0 y 1. Las técnicas de soft computing, basadas en la lógica difusa, pueden ser útiles aquí. Mucho más cerca de la forma en que funciona el cerebro humano agregando datos a verdades parciales, esta lógica es uno de los principales aspectos exclusivos de CI.

Dentro de los mismos principios de lógica difusa y binaria siguen los sistemas crispy y fuzzy. La lógica nítida (crisp) es parte de los principios de la inteligencia artificial y consiste en incluir o no un elemento en un conjunto, mientras que los sistemas difusos permiten que los elementos estén parcialmente en un conjunto. Siguiendo esta lógica, a cada elemento se le puede dar un grado de pertenencia (de 0 a 1) y no exclusivamente uno de estos 2 valores.

Sistemas Difusos

La lógica clásica solo permite conclusiones que son verdaderas o falsas. Sin embargo, también hay proposiciones con respuestas variables, como las que se pueden encontrar al pedir a un grupo de personas que identifiquen un color. En tales casos, la verdad aparece como resultado de un razonamiento a partir de un conocimiento inexacto o parcial en el que las respuestas muestreadas se mapean en un espectro.

Tanto los grados de verdad como las probabilidades oscilan entre 0 y 1 y, por lo tanto, pueden parecer similares al principio, pero la lógica difusa usa grados de verdad como un modelo matemático de vaguedad, mientras que la probabilidad indica cuanto más probable es que ocurra un evento.

Utilizando el lenguaje humano como fuente de inspiración, los sistemas difusos (FS) modelan la imprecisión lingüística y resuelven problemas inciertos basados en una generalización de la lógica tradicional, lo que nos permite realizar razonamientos aproximados. Esta área de investigación incluye conjuntos y sistemas difusos, agrupamiento y clasificación difusa, controladores difusos, resumen lingüístico, redes neuronales difusas, conjuntos y sistemas difusos de tipo 2, etc.

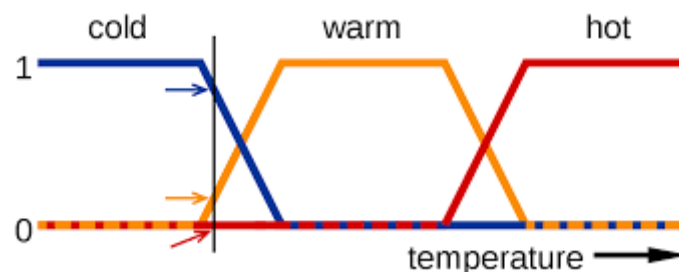
El término lógica difusa se introdujo con la propuesta de la [teoría de conjuntos difusos](#) en 1965 del matemático [Lotfi A. Zadeh](#). En la Antigua Grecia, Aristóteles había planteado que la lógica bivalente no era aplicable a sucesos futuros, como por ejemplo la declaración "Habrà una batalla naval mañana". En otras palabras, los enunciados sobre el futuro no eran ni verdaderos ni falsos, pero

se les podía asignar un valor intermedio, para representar su posibilidad de convertirse en verdaderos en el futuro. En la primera parte del siglo XX, [J. Lukasiewicz](#) y [A. Tarski](#) habían publicado trabajos sobre lógica multivaluada.

La lógica difusa se basa en la observación de que las personas toman decisiones basadas en información imprecisa y no numérica. Los conjuntos borrosos son medios matemáticos para representar la vaguedad y la información imprecisa (de ahí el término borroso). Estos modelos tienen la capacidad de reconocer, representar, manipular, interpretar y utilizar datos e información imprecisos y poco seguros.

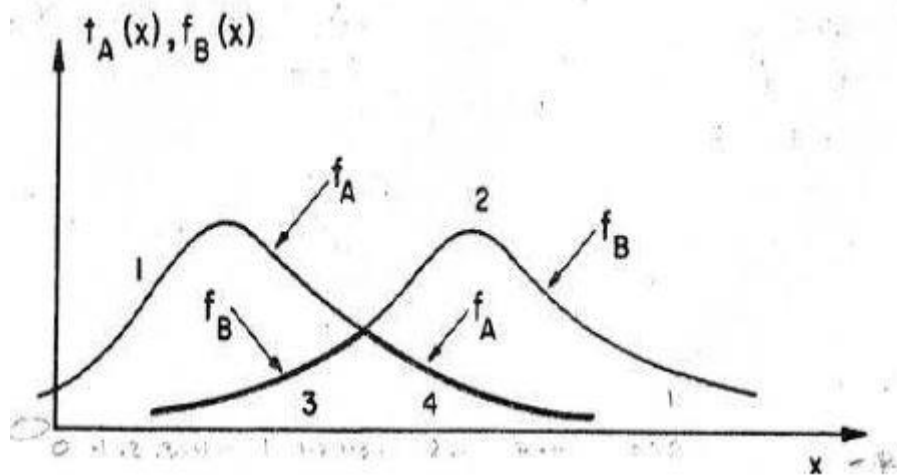
La lógica difusa, uno de los principios fundamentales de CI, consiste en mediciones y modelado de procesos realizados para procesos complejos de la vida real. Puede enfrentarse a datos incompletos y, lo que es más importante, a la ignorancia de los datos en un modelo de proceso, a diferencia de la inteligencia artificial, que requiere un conocimiento exacto.

Una aplicación básica podría caracterizar varios sub-rangos de una variable continua. Por ejemplo, una medición de temperatura para frenos antibloqueo puede tener varias funciones de membresía separadas que definen rangos de temperatura particulares necesarios para controlar los frenos correctamente. Cada función asigna el mismo valor de temperatura a un valor real en el rango de 0 a 1. Estos valores de verdad se pueden usar para determinar cómo se deben controlar los frenos. La teoría de conjuntos borrosos proporciona un medio para representar la incertidumbre.



Una variable lingüística como la edad puede aceptar valores como joven y su antónimo viejo. Debido a que los lenguajes naturales no siempre contienen suficientes términos de valor para expresar una escala de valores difusos, es una práctica común modificar los valores lingüísticos con adjetivos o adverbios. Por ejemplo, podemos usar las palabras bastante y algo para construir los valores adicionales bastante viejo o algo joven.

Se trata de un tipo de sistema lógico computacional dirigido a la categorización de datos empíricos de forma cercana a la empleada por el razonamiento humano. Esto significa que, en vez de clasificar utilizando el modelo binario (0, 1), recoge los matices y las calificaciones intermedias. Se entiende, en cierto modo, como la valorización del espacio entre lo verdadero y lo falso. La dificultad reside en la voluntad de captar la falta de precisión de la lengua humana y hacerla comprensible por una computadora.



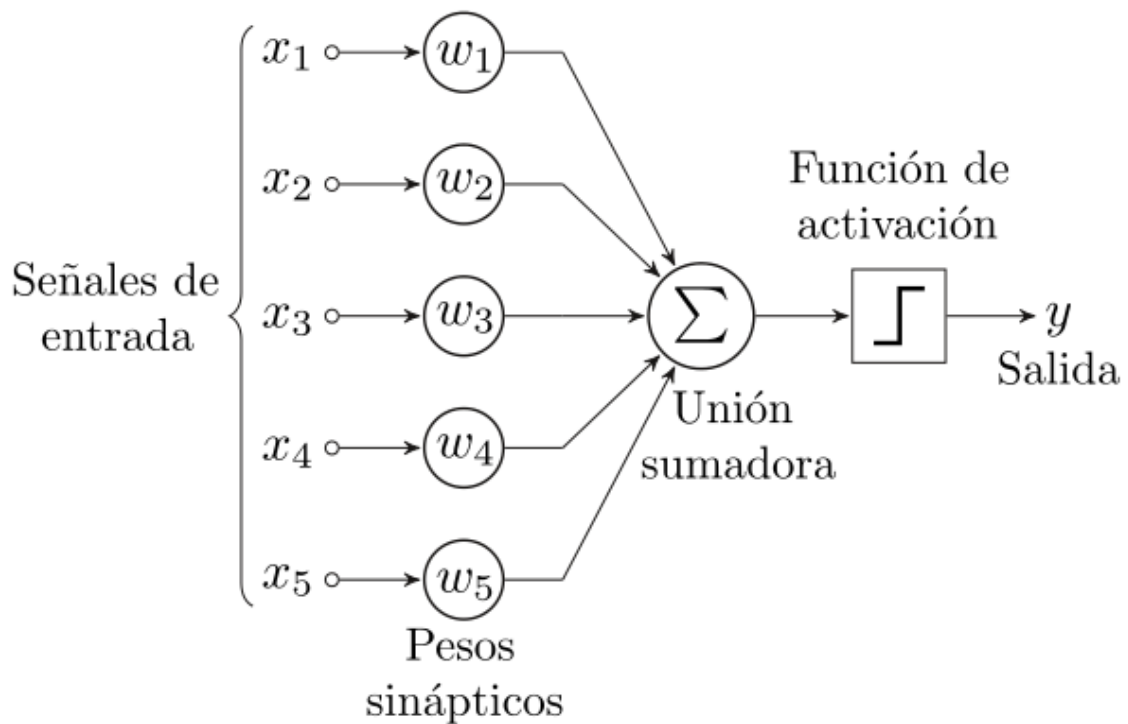
Esta técnica tiende a aplicarse a una amplia gama de dominios como el control, el procesamiento de imágenes y la toma de decisiones, así como en el campo de los electrodomésticos con lavadoras, hornos microondas, etc. También se emplea cuando utilizamos una cámara de vídeo, donde ayuda a estabilizar la imagen mientras sujetamos la cámara de forma inestable.

Redes Neuronales

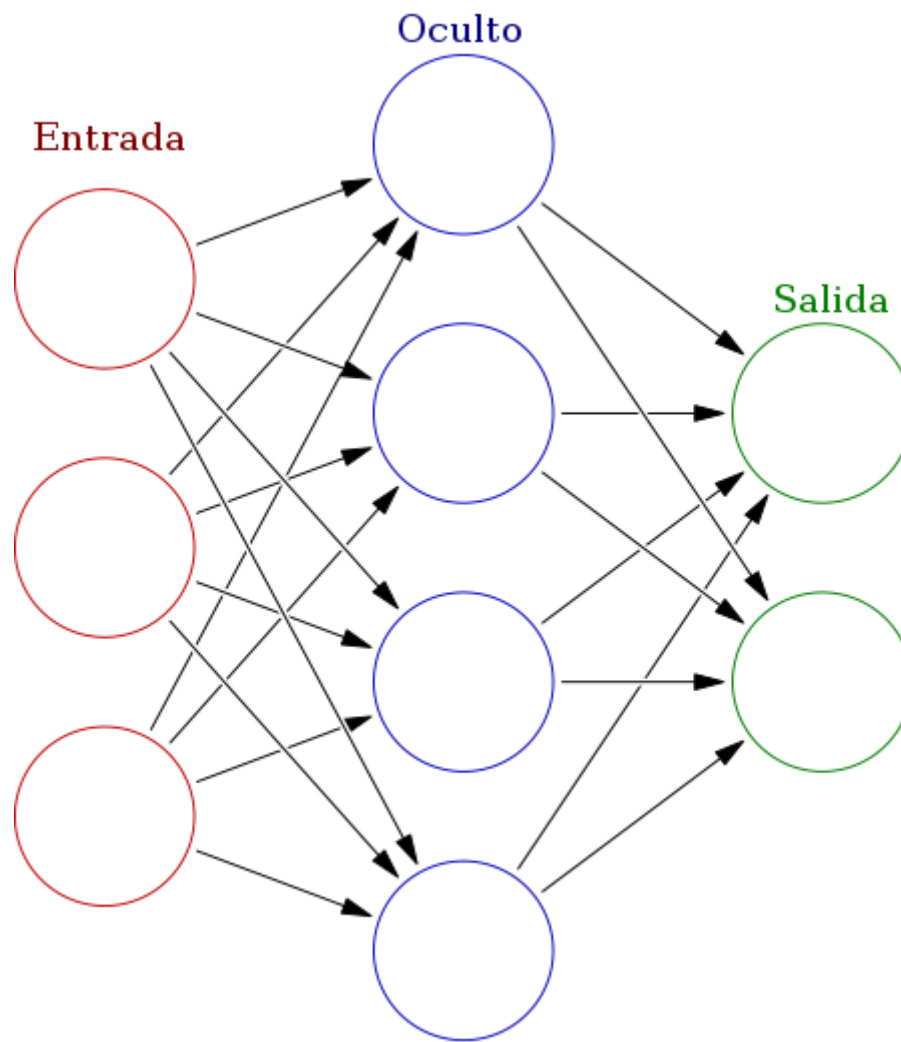
Utilizando el cerebro humano como fuente de inspiración, las redes neuronales artificiales (NN) son redes distribuidas masivamente en paralelo que tienen la capacidad de aprender y generalizar a partir de ejemplos. Esta área de investigación incluye redes neuronales retroalimentadas, redes neuronales recurrentes, redes neuronales autoorganizadas, aprendizaje profundo, redes neuronales convolucionales, etc.

En 1925 [Wilhelm Lenz](#) y [Ernst Ising](#) crearon y estudiaron el [modelo Ising](#), el cual es en esencia un [red neuronal recurrente](#) (RNN) artificial sin capacidad de aprendizaje, la cual consistía en elementos de umbral similares a neuronas. Posteriormente, en 1972, [Shun'ichi Amari](#) introdujo una versión con capacidad de aprendizaje, la cual fue popularizada por [John Hopfield](#) en 1982. En 1958, [Frank Rosenblatt](#) inventó el [perceptron](#), la primera red neuronal artificial implementada.

Se trata de modelos matemáticos artificiales inspirados por el funcionamiento del cerebro humano. Buscan imitarlo con el fin de crear un sistema inteligente que actúe y responda de forma parecida a como lo harían las neuronas. Se establece un sistema de neuronas conectadas con una memoria local limitada que se interconectan mediante canales de comunicación (o conexiones) que, usualmente, transmiten datos numéricos codificados y pueden tener fuerzas variables (o ponderaciones). Cada neurona tendrá un peso determinado que, junto con la cantidad de entradas que reciba, determinará su fuerza de actuación.



Las diferentes neuronas se juntan en grupos que, a su vez, se clasificarán por capas. La primera capa (que no tiene ninguna anterior que le suministre la información) se conoce como capa de entrada, la última como capa de salida (resultado visible de la red), y todas aquellas que se encuentran en medio reciben el nombre de capas ocultas. Puede no haber ninguna capa oculta, pero para conformar una red se requiere una capa de entrada y una de salida (en caso de sólo haber una capa, ésta efectuará los dos roles a la vez).



La red aprende revisando registros individuales, realizando predicciones por cada uno de ellos y adaptándose cuando se equivoca. Este proceso se lleva a cabo multitud de ocasiones hasta que el sistema llega a un punto de conclusión. Inicialmente, todas las ponderaciones son aleatorias y las respuestas incorrectas. La red, entonces, va aprendiendo a medida que experimenta.

En 1992 se publicó una alternativa a las RNN que ahora se llama Transformador lineal o Transformador con autoatención linealizada. Aprende focos internos de atención: una red neuronal de retroalimentación lenta aprende por descenso de gradiente a controlar los pesos rápidos de otra red neuronal a través de productos externos de patrones de activación autogenerados DESDE y HASTA (actualmente se denominan clave y valor para el mecanismo de atención). Este mapeo de atención se aplica a un patrón de consulta.

El concepto de Transformador fue introducido en el artículo de 2017, "[Attention Is All You Need](#)". Los transformadores se han convertido cada vez más en el modelo elegido para el procesamiento del lenguaje natural. Muchos modelos

modernos de lenguaje grande, como ChatGPT, GPT-4 y BERT, lo utilizan. Los transformadores también se utilizan cada vez más en la visión artificial.

Por lo tanto, las redes neuronales artificiales están dotadas de sistemas de procesamiento de información distribuida, lo que permite el proceso y el aprendizaje a partir de datos experimentales. Trabajando como seres humanos, la tolerancia a fallas es también uno de los principales activos de este principio.

En cuanto a sus aplicaciones, las redes neuronales se pueden clasificar en cinco grupos: análisis y clasificación de datos, memoria asociativa, generación de patrones en clúster y control. Generalmente, este método tiene como objetivo analizar y clasificar datos médicos, proceder a la detección de rostros y fraudes, y lo más importante, lidiar con las no linealidades de un sistema para controlarlo. Además, las técnicas de redes neuronales comparten con las de lógica difusa la ventaja de permitir la agrupación de datos.

Cómputo Evolutivo

Utilizando la evolución biológica como fuente de inspiración, el cómputo evolutivo (EC) resuelve problemas de optimización generando, evaluando y modificando una población de posibles soluciones. EC incluye algoritmos genéticos, programación evolutiva, estrategias de evolución, programación genética, inteligencia de enjambre, evolución diferencial, hardware evolutivo, optimización multiobjetivo, etc.

Basado en el proceso de selección natural, el cómputo evolutivo consiste en capitalizar la fuerza de la evolución natural para generar nuevas metodologías evolutivas artificiales. También incluye otras áreas como la estrategia de evolución y los algoritmos evolutivos que son vistos como solucionadores de problemas. Se trata de técnicas computacionales para resolver problemas principalmente de optimización combinatoria. Requiere de la heurística y basa su operación en los principios de la evolución de Darwin.

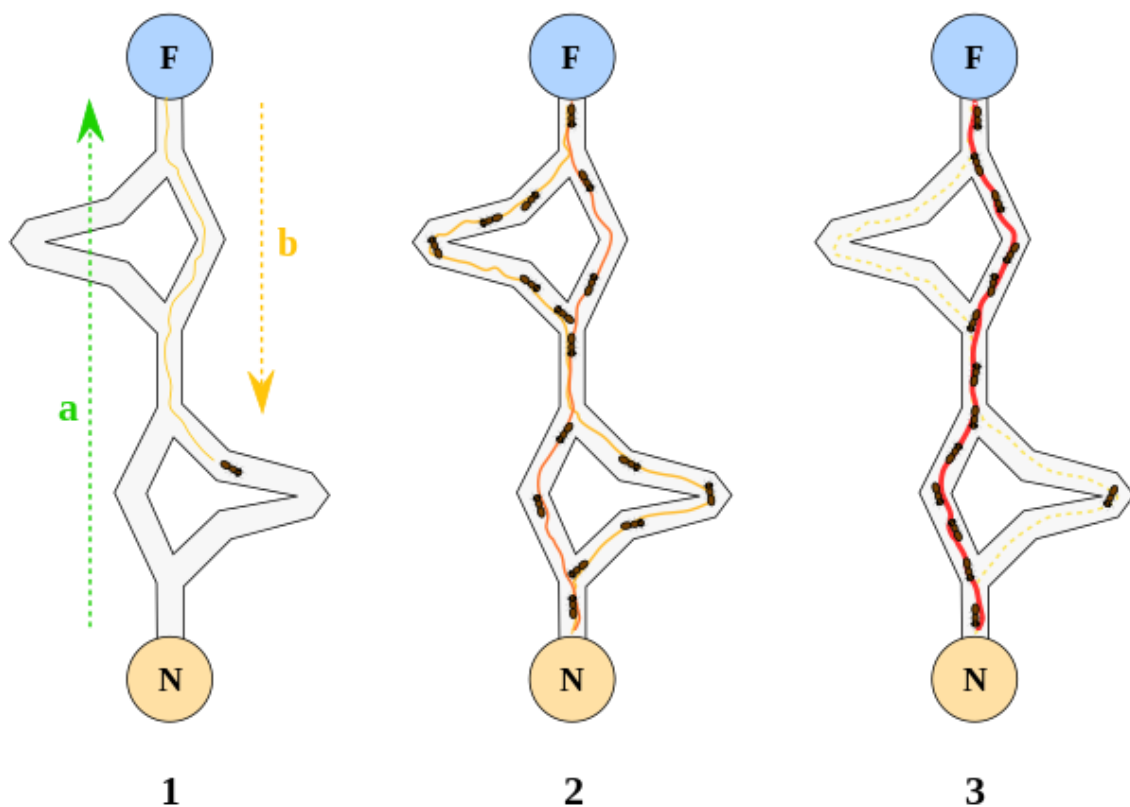
Existen muchas definiciones para la inteligencia de enjambre o swarm intelligence (SI). Básicamente, se trata de una rama de la inteligencia artificial inspirada en la naturaleza, concretamente en sistemas biológicos en los que pueden observarse comportamientos colectivos.

La inteligencia de enjambre describe el comportamiento colectivo de sistemas descentralizados y autoorganizados, naturales o artificiales. Los sistemas SI se componen típicamente de una población de agentes simples que interactúan localmente entre sí y con su entorno.

Los agentes siguen reglas muy simples. No existe una estructura de control centralizada que dicte cómo deben comportarse los agentes. Las interacciones locales y aleatorias entre dichos agentes conducen a la aparición de un comportamiento global "inteligente", desconocido para los agentes individuales. El comportamiento de las hormigas fue fuente de inspiración de una técnica de optimización basada en metaheurísticas denominada ACO (Ant Colony Optimization)



Los algoritmos de optimización de colonias de hormigas (ACO) modelan las acciones de una colonia de hormigas. Los ACO son útiles para encontrar caminos hacia los objetivos mediante optimizaciones metaheurísticas. Las 'hormigas' artificiales localizan soluciones óptimas moviéndose a través de un espacio de parámetros que representa todas las soluciones posibles. En el mundo natural, las hormigas inicialmente deambulan al azar. Al encontrar comida, regresan a su colonia mientras dejan rastros de feromonas. Si otras hormigas encuentran ese camino, es probable que no sigan viajando al azar, sino que sigan el rastro, regresando y reforzándolo si finalmente encuentran comida.

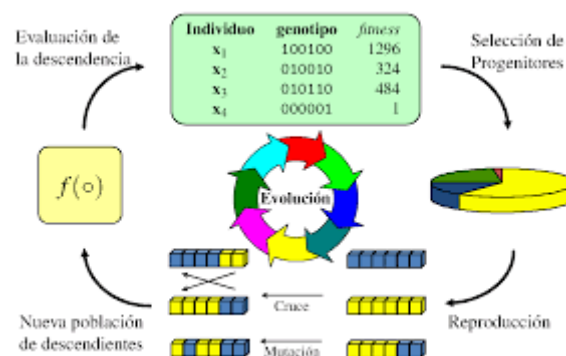


El rastro de feromonas comienza a evaporarse con el tiempo, reduciendo así su fuerza de atracción. Cuanto más tiempo tarda una hormiga en viajar hacia abajo y hacia atrás, más tiempo se evaporan las feromonas. Por lo tanto, la densidad de feromonas aumenta en los trayectos más cortos que en los más largos.

El algoritmo de la colonia de hormigas imita este comportamiento con "hormigas simuladas" caminando alrededor del gráfico que representa el

problema a resolver. Las 'hormigas' simuladas registran de manera similar sus posiciones y la calidad de sus soluciones, de modo que en posteriores iteraciones de simulación más hormigas localicen mejores soluciones.

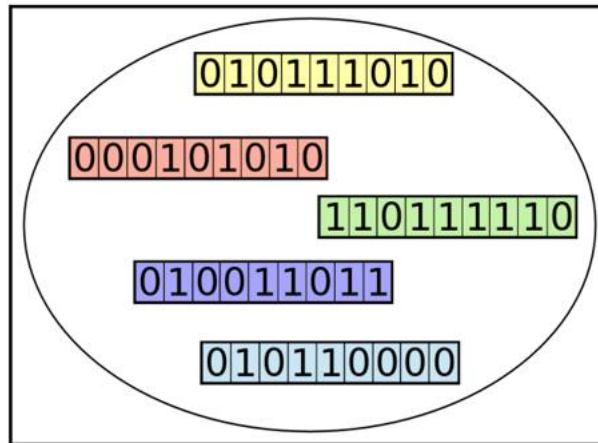
Para trabajar con casos hipotéticos, la computación evolutiva (CE) hace uso de una metáfora llamada metáfora evolutiva. Se refiere como población a un grupo de individuos que coexisten en un entorno concreto valiéndose de recursos limitados. La competición por estos recursos es lo que desemboca en una selección de los más fuertes capaces de sobrevivir, los mejor adaptados. Estos sujetos pasarán a ser los progenitores de las nuevas generaciones de individuos a los que darán lugar mediante mutaciones y recombinaciones (crossovers). Se vuelve a repetir la rueda, y estos nuevos individuos compiten, de nuevo, por la supervivencia, mejorando así, progresivamente, la calidad de los que sobreviven.



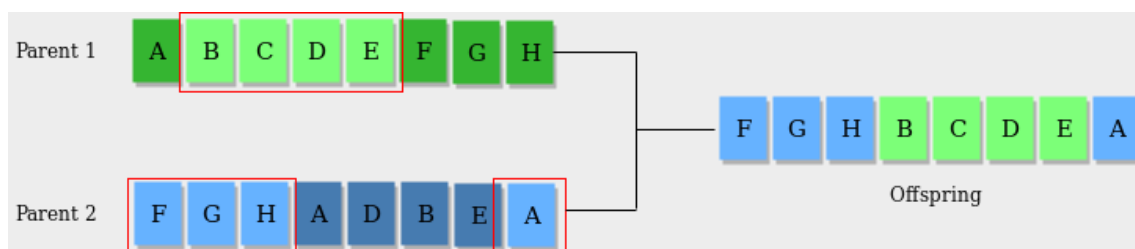
Un algoritmo genético (Genetic Algorithm GA) es un algoritmo de optimización evolutiva inspirado en la selección natural biológica con una representación simple y operadores basados en recombinación genética y mutaciones. GA utiliza metaheurística para encontrar una solución global estocástica a un problema. Es uno de los algoritmos de inspiración biológica ampliamente conocidos.

Los GA simulan el proceso de adaptación de las especies a los cambios en su entorno para sobrevivir y reproducirse, es decir, la "supervivencia del más apto" entre individuos de generaciones consecutivas para resolver un problema. Cada generación consiste en una población de individuos que representan puntos en el espacio de búsqueda y posibles soluciones.

El algoritmo crea primero una población de cadenas de bits aleatorias de tamaño fijo. El bucle principal se repite durante varias iteraciones o hasta que no haya más mejoras. En cada iteración, los individuos se evalúan en función de sus puntuaciones de aptitud física. Una iteración representa una generación evolutiva.



Los padres se toman en parejas y se recombinan para crear dos hijos.



La mutación es un cambio aleatorio en un cromosoma para introducir nuevos patrones. La mutación implica cambiar bits en las soluciones creadas para los descendientes.



Las principales aplicaciones de este principio cubren áreas como la optimización y la optimización multiobjetivo, a las que las técnicas matemáticas tradicionales ya no son suficientes para aplicar a una amplia gama de problemas como análisis de ADN, problemas de programación, etc.

Teoría del Aprendizaje Computacional

Buscando una forma de "razonar" cercana a la humana, la teoría del aprendizaje es uno de los principales enfoques de la IC. En informática, la teoría del aprendizaje computacional (o simplemente la teoría del aprendizaje) es un subcampo de la inteligencia artificial dedicado al estudio del diseño y análisis de algoritmos de aprendizaje automático.

En psicología, el aprendizaje es el proceso de reunir efectos y experiencias cognitivas, emocionales y ambientales para adquirir, mejorar o cambiar conocimientos, habilidades, valores y visiones del mundo. Aprender teorías ayuda a comprender cómo se procesan estos efectos y experiencias, y luego ayuda a hacer predicciones basadas en experiencias previas.

La teoría del aprendizaje computacional, o CoLT para abreviar, es un campo de estudio relacionado con el uso de métodos matemáticos formales aplicados a los sistemas de aprendizaje. Se busca utilizar las herramientas de la informática teórica para cuantificar problemas de aprendizaje. Esto incluye caracterizar la dificultad de aprender tareas específicas.

La teoría del aprendizaje computacional puede considerarse como una extensión o un hermano de la teoría del aprendizaje estadístico, o SLT para abreviar, que utiliza métodos formales para cuantificar los algoritmos de aprendizaje.

El enfoque en la teoría del aprendizaje computacional suele estar en las tareas de aprendizaje supervisado. El análisis formal de problemas reales y algoritmos reales es muy desafiante. Como tal, es común reducir la complejidad del análisis centrándose en tareas de clasificación binaria e incluso en sistemas binarios simples basados en reglas. Como tal, la aplicación práctica de los teoremas puede ser limitada o difícil de interpretar para problemas y algoritmos reales.

Los resultados teóricos en el aprendizaje automático se ocupan principalmente de un tipo de aprendizaje inductivo llamado aprendizaje supervisado. En el aprendizaje supervisado, un algoritmo recibe muestras que están etiquetadas de alguna manera útil. Por ejemplo, las muestras pueden ser descripciones de hongos y las etiquetas pueden indicar si los hongos son comestibles o no. El algoritmo toma estas muestras previamente etiquetadas y las usa para inducir un clasificador. Este clasificador es una función que asigna etiquetas a las muestras, incluidas las muestras que el algoritmo no ha visto previamente. El objetivo del algoritmo de aprendizaje supervisado es optimizar alguna medida de rendimiento, como minimizar la cantidad de errores cometidos en nuevas muestras.

La principal pregunta sin respuesta en el aprendizaje es esta: ¿Cómo podemos estar seguros de que nuestro algoritmo de aprendizaje ha producido una hipótesis que predecirá el valor correcto para entradas no vistas previamente?

Las preguntas exploradas en la teoría del aprendizaje computacional pueden incluir:

- ¿Cómo sabemos que un modelo tiene una buena aproximación para la función objetivo?
- ¿Qué espacio de hipótesis se debe utilizar?
- ¿Cómo sabemos si tenemos una buena solución local o global?
- ¿Cómo evitamos el sobreajuste?
- ¿Cuántos ejemplos de datos se necesitan?

Existen diversos enfoques para la teoría del aprendizaje computacional basados en asumir diferentes suposiciones sobre los principios de inferencia utilizados para generalizar a partir de datos limitados. Esto incluye diferentes

definiciones de probabilidad como probabilidad de frecuencia, probabilidad bayesiana, etc., y diferentes suposiciones sobre la generación de muestras.

Métodos Probabilísticos

Las dos principales fuentes de incertidumbre son la aleatoriedad, o incertidumbre probabilística, y la borrosidad, que encarna la imprecisión por vaguedad o falta de conocimiento, y son complementarias. Los modelos que integran ambos enfoques se conocen como sistemas borrosos probabilísticos (PFS). Esencialmente, el PFS es una metodología que se basa en un sistema de inferencia difusa (FIS), que se ha modificado para adaptarse a una base de reglas difusas probabilísticas. Es decir, se aplica un marco probabilístico a un modelo difuso existente, bajo el cual se tratan simultáneamente las incertidumbres difusa y probabilística. Esto proporciona un mapeo estocástico de entrada-salida entre los conjuntos borrosos de entrada asociados con la parte antecedente de la base de reglas y los conjuntos borrosos de salida asociados con la parte consecuente.

Los métodos probabilísticos tienen como objetivo evaluar los resultados de un sistema Computacional Inteligente, definido principalmente por la aleatoriedad. Por lo tanto, los métodos probabilísticos sacan a relucir las posibles soluciones a un problema, en base al conocimiento previo y son uno de los elementos principales de la lógica difusa.

Análisis de datos guiado por Inteligencia Computacional

Un camino para combinar IC y ML¹

Las tendencias actuales de la sociedad contrastan con la **percepción común de que la gente está cada vez más educada**. Sin embargo, los hechos demuestran que vivimos en una época en la que se observa:

Una “innumerabilidad” rampante (es decir, **ignorancia en temas relacionados con las matemáticas**, incluyendo conceptos muy básicos) en la población media, independientemente de la edad, el género y la clase social. Como ejemplo, véase este caso:
"¿Por qué debemos pagar la misma cantidad por un tercio de libra de carne que por un cuarto de libra de carne en McDonald's?
Nos están cobrando de más".

Las tendencias actuales de la sociedad contrastan con la **percepción común de que la gente está cada vez más educada**. De hecho, los hechos demuestran que vivimos en una época en la que se observa:

- Una disminución general de la conexión con la realidad y el conocimiento, que termina en una ignorancia auto justificada de los hechos (¡porque puedes confiar en Google!).
- La incredulidad ampliamente aceptada de que existe (y puede existir) una realidad objetiva real.

¹ Keynote de Doug Lenat , CEO of Cycorp , Inc. en el AAAI MAKE 2021 , Stanford University, Palo Alto, California, USA.

- La suposición de que las respuestas a las preguntas están siempre disponibles...

Hoy en día, la gente confía en las máquinas para obtener conocimientos, mientras que los sistemas de inferencia actuales **no tienen la lógica necesaria** para reunir conocimientos de muchas fuentes diferentes y sacar conclusiones o respuestas.

Por ejemplo, ¿es la catedral de Ciudad Real más alta que la torre del Micalet? La respuesta no se puede encontrar de forma completamente automática en Google, porque ninguna página web existente habla de ello explícitamente, aunque **las piezas individuales de la respuesta** están disponibles...

- (Question Answering Systems...)

Esto se debe a las dificultades del aprendizaje automático:

- Situaciones en las que los expertos no están de acuerdo con los resultados
- Riesgo de sesgo de confirmación
- Respuestas absurdas o conclusiones intermedias
- Poca cantidad de datos disponibles
- Cuando la respuesta implica el "sentido común"
- Cuando los mejores expertos son mucho mejores que la persona media

Hay cosas que el aprendizaje automático puede hacer sobre **textos generados por humanos** (desde las redes sociales al Quijote) para entender preguntas y respuestas, utilizando diferentes lógicas limitadas: desde la **lógica proposicional** hasta la **lógica completa de primer orden**.

Pero la cantidad de preguntas que un humano puede formular y que podemos responder es **extremadamente limitada** (aún hoy en día) con respecto a las preguntas que cualquier humano que lea un texto o vea una representación argumental o una película puede responder. Y, sobre todo, hay una pregunta que la IA apenas puede responder: ¿**porqué**?

La mayor parte de las veces, la lógica de primer orden no es **suficientemente rápida a nivel computacional** (por no hablar de la lógica de segundo orden o de orden superior).

La idea propuesta por Lenat² consiste en dividir los **dos ingredientes importantes del razonamiento**, para hacer frente al **compromiso entre expresividad y eficiencia**:

- aspectos epistemológicos basados en el conocimiento científico
- aspectos heurísticos

Al separar los dos, se pueden utilizar lenguajes extremadamente **eficientes** para la parte **heurística**, mientras se reservan lenguajes **muy potentes** (y menos eficientes) para la parte **epistemológica**.

Tenemos **dos lenguajes en el juego**: el lenguaje de nivel heurístico (HL) y el lenguaje de nivel epistemológico (EL).

² Keynote de Doug Lenat, CEO of Cycorp, Inc. en el AAAI MAKE 2021, Stanford University, Palo Alto, California, USA.

La otra aportación interesante al razonamiento es el **contexto**: el conocimiento **puede ser diferente en distintos contextos**.

El sistema es capaz de utilizar la **ABDUCCIÓN** (o razonamiento abductivo) en contraste con la **DEDUCCIÓN** para proponer posibilidades probables en las respuestas como, por ejemplo, por qué ocurre algo en una historia. En el caso de

- una **deducción** se obtiene una [conclusión](#) «q» de una [premisa](#) «p», mientras que
- el razonamiento abductivo consiste en explicar «q» mediante «p» considerando p como hipótesis explicativa. En la abducción, a fin de entender un fenómeno se introduce una regla que opera en forma de [hipótesis](#) para considerar dentro de tal regla al posible resultado como un caso particular. El **razonamiento abductivo** es un tipo de [razonamiento](#) que, a partir de la descripción de un [hecho](#) o [fenómeno](#), ofrece o llega a una [hipótesis](#) que explica las posibles razones o motivos del hecho mediante las [premisas](#) obtenidas.

El razonamiento abductivo (también llamado abducción, inferencia abductiva o retroducción) es una forma de inferencia lógica que busca la conclusión más simple y probable de un conjunto de observaciones. Fue formulado y propuesto por el filósofo estadounidense [Charles Sanders Peirce](#) a partir del último tercio del siglo XIX.

El razonamiento abductivo, a diferencia del razonamiento deductivo, produce una conclusión plausible pero no la verifica definitivamente. Las conclusiones abductivas no eliminan la incertidumbre o la duda, que se expresa en términos como "mejor disponible" o "más probable". Uno puede entender el razonamiento abductivo como inferencia a la mejor explicación, aunque no todos los usos de los términos abducción e inferencia a la mejor explicación son equivalentes.

El razonamiento abductivo permite inferir 'a' como explicación de 'b'. Como resultado de esta inferencia, la abducción permite que la condición previa 'a' sea abducida de la consecuencia 'b'. El razonamiento deductivo y el razonamiento abductivo difieren así en qué extremo, izquierdo o derecho, de la proposición "a implica b" sirve como conclusión.

Por ejemplo, en un juego de billar, después de mirar y ver la bola ocho moviéndose hacia nosotros, podemos abducir que la bola blanca golpeó la bola ocho. El golpe de la bola blanca explicaría el movimiento de la bola ocho. Sirve como una hipótesis que explica nuestra observación. Dadas las muchas explicaciones posibles para el movimiento de la bola ocho, nuestra abducción no nos deja seguros de que la bola blanca de hecho golpeó la bola ocho, pero nuestra abducción, aún útil, puede servir para orientarnos en nuestro entorno.

A pesar de las muchas explicaciones posibles para cualquier proceso físico que observamos, tendemos a abducir una sola explicación (o algunas explicaciones) para este proceso con la expectativa de poder orientarnos mejor en nuestro entorno y descartar algunas posibilidades. El razonamiento

abductivo, correctamente utilizado, puede ser una fuente útil de antecedentes en las estadísticas bayesianas.

RAZONAMIENTO APROXIMADO: **Lógica Borrosa**.

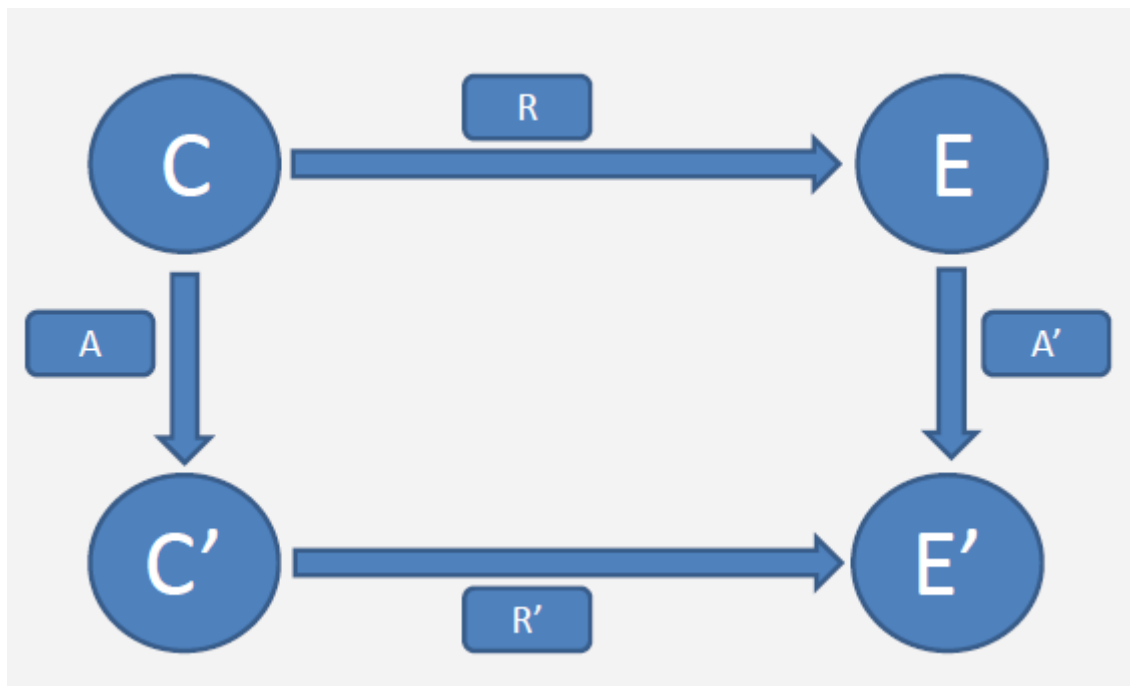
Por lo tanto, el camino hacia la **combinación del aprendizaje automático y las bases de conocimiento** consiste en explotar su sinergia latente. En la práctica:

- Usar las bases de conocimiento como **generadoras de** una gran cantidad de **datos de entrenamiento** para los sistemas de aprendizaje automático.
- Utilizar el aprendizaje automático para **generar hipótesis** y, a continuación, utilizar las bases de conocimiento para **comprobar si hay errores o contradicciones**.

Uno de los principales efectos a largo plazo es la posibilidad de trasladar el poder de análisis **de la correlación a la causalidad**.

Esto puede aportar un enfoque **totalmente nuevo** a todo el campo.

De la correlación a la causalidad...



Bibliografía

- [Inteligencia computacional](#)
- [Computational intelligence](#)
- [What is Computational Intelligence?](#)
- [A Gentle Introduction to Computational Learning Theory](#)
- [An Overview of Probabilistic Fuzzy Systems -- Some Preliminary Observations](#) 2017
- [IEEE Computational Intelligence Society History](#)

- Soft computing y análisis inteligente de datos, José Ángel Olivas Varela
- [Abductive reasoning](#)