La evolución de las máquinas ante los grandes retos que enfrenta la sociedad

Juan José Arango Serrano
Escuela de Ingeniería de Sistemas e Informática
Universidad Industrial de Santander
Bucaramanga, Colombia
jjarangoserrano@gmail.com

María Camila Aparicio Nova
Escuela de Ingeniería de Sistemas e Informática
Universidad Industrial de Santander
Bucaramanga, Colombia
mcamilaanova@gmail.com

Bryan Daniel Alvarez Castillo
Escuela de Ingeniería de Sistemas e Informática
Universidad Industrial de Santander
Bucaramanga, Colombia
bryanalvarezc1998@gmail.com

Carlos Javier Aceros Blanco
Escuela de Ingeniería de Sistemas e Informática
Universidad Industrial de Santander
Bucaramanga, Colombia
hola@carlosaceros.co

Resumen

La inteligencia artificial se ha convertido en una de las tecnologías avanzadas fundamentada en la inteligencia y autoaprendizaje de las máquinas con base a la razón y la conducta, con destreza para enfrentar grandes retos de la sociedad actual. Recientemente ha sido partícipe en la revolución del campo de los negocios, la medicina, la visión, el habla, el entendimiento del lenguaje, entre otros. Incluso ha venido superando la precisión humana, sin embargo, su eficiencia se debe al trabajo que por dentro de sí misma lleva, la cual se compone de un álgebra lineal de baja precisión. En cuanto al hardware se refiere, su rendimiento depende del sistema computacional en el que esté diseñado y proporcionado por los propios expertos en el área y quienes trabajan de la mano de la arquitectura computacional.

Palabras Clave

Inteligencia Artificial, Tecnología, Machine Learning, Computación.

Abstract

Artificial Intelligence (AI) has become one of the most advanced technologies in the self-learning and intelligence of machines based on reasoning and machine behavior, with skill to face the great challenges of today's society. It has been part of the revolution in the field of business, medicine, vision, speech, understanding of language, etc. in fact AI has even surpassed human precision. In terms of hardware, its performance depends on the computer system where it was designed, provided by experts in the field either computer engineers or computer scientist and computer architects.

Keywords

Artificial Intelligence, Technology, Machine Learning, Computing.

I. Introducción

A través del paso de los años, la humanidad ha demostrado su capacidad para prolongar su existencia y el desarrollo de la civilización ha dependido completamente de las herramientas que el ser humano es capaz de crear, inventar y mejorar. Desde tiempos remotos, ha sido notoria la importancia de la tecnología en el desarrollo humano, especialmente en los últimos siglos.

En la actualidad, la civilización moderna afronta diversos retos en distintos campos y nace de esta problemática la necesidad de poder atender nuestros problemas y de allí, el nacimiento de las máquinas, herramienta esencial en la vida del hombre actual y con esto, el surgimiento de las computadoras, máquinas con capacidad de realizar diversas tareas las cuales le tomarían al hombre hasta inclusive años en realizar como lo son los enormes cálculos matemáticos, predicciones del clima o el comportamiento de cierto mercado, la precisión de lanzamiento de un cohete, entre muchas otras. Con esto, ha venido tomando importancia el mayor reto tecnológico de la historia, el desarrollo de la inteligencia artificial la cual ha tenido un incremento exponencial en las últimas décadas y ha sido papel importante en el afronte de los problemas de la sociedad actual a tal punto en el que incluso el último de los grandes genios de la humanidad, Stephen Hawking, ha cuestionado el poder de la inteligencia y autoaprendizaje de las máquinas.

El centro de este artículo radica en la inteligencia artificial (IA), especialmente en su evolución a través de los años y una comparativa de la percepción de esta tecnología en las que en un principio no fue más que un concepto ideal sin una puesta en práctica tan efectiva como lo vemos hoy en día y que, de hecho, sigue creciendo y generando cada vez nuevas expectativas en la sociedad actual.

II. Desarrollo

La IA se ha especulado desde tiempos atrás, principalmente en la imaginación de filósofos, o escritores, sin embargo una de las primeras máquinas que se podían catalogar como IA fue la máquina de Turing presentada por el inglés Alan Turing, dicha maquina era un prototipo de un autómata infinito, el cual seguía una serie de instrucciones por si misma hasta llegar a un resultado, este concepto fue la raíz para la computación actual, esta fue mejorada y utilizada en la segunda guerra mundial con el fin de codificar y mandar mensajes a los aliados. Anteriormente, los gobiernos mantenían oculta gran parte de los datos, sin embargo, actualmente estos datos son fácilmente accesibles a la comunidad en general, lo cual hace que actualmente se tengan bastantes datos, sobre las personas en general, y nos permiten poder trabajar en ellos, desarrollando así técnicas para tratarlos de manera eficiente. En Colombia la IA no ha incursionado del todo dentro del país por diversos tipos de obstáculos que se encuentra en el camino, como la falta de inversión en ciencia y tecnología que se hace por el gobierno colombiano y por el imaginario colectivo de las personas, que no cree que sea útil y no conocen la gama tan amplia que hay en sus usos. A pesar de todo, cabe destacar que Medellín es la ciudad colombiana que más le apuesta a la innovación y es tanto así que utiliza el de su PIB para este renglón.

En el año 2018 creó un convenio con el Instituto para la Automatización Robótica de Procesos e Inteligencia Artificial (IRPA AI) de Estados Unidos con el que conjuntamente se creará un Centro de Excelencia de esta disciplina. Este Centro generará desde Medellín, servicios avanzados en automatización de procesos e IA, inicialmente las personas que sean allí formadas utilizaran sus conocimientos en granes empresas dedicadas a BPO (Business Process Outsourcing). Al mismo tiempo, con este proyecto se espera crear un impacto dentro del país y activar la demanda dentro de este, el representante de la alcaldía de Medellín Alejandro Franco perteneciente al grupo N que lidera esta iniciativa cree posible una integración de la IA a una flota de vehículos eléctricos dentro del sistema de transporte masivo de la ciudad y propone que este servicio a futuro cubra el 20 por ciento de los desplazamientos en Medellín.

Otro ejemplo importante del impacto positivo que puede llegar a tener la implementación de la IA en Colombia lo da Andrés Jiménez director de operaciones de Microsoft en Colombia, que explicó que junto a Globo estudio se desarrolló un programa que permite tomarle foto a los capullos de las flores para clasificar el estado fenológico (o de desarrollo) y que determina la fecha en la que se pueden arrancar las flores, lo que permite a los floricultores planificarse para fechas comerciales y aumentar la productividad. Estos ejemplos son demostraciones contundentes que la inteligencia artificial facilita la vida de las personas y optimiza la productividad de las empresas y a pesar de lo escasa que es la utilización de la IA en nuestro país es algo que con el tiempo necesitaremos.

III. Arquitectura en la Inteligencia Artificial

Las aplicaciones de IA abarcan todo el espectro de diseño de hardware, desde el centro de datos y las supercomputadoras, hasta aplicaciones automotrices, dispositivos móviles, entre otro. Estos puntos de diseño vienen con un paquete de restricciones de costo, potencia, tamaño, y mezcla de computo, comunicación y almacenamiento de datos requeridos para una solución, como se menciona en el artículo de Gordon Bell y Peter Freeman [1] la maquina definitiva esta aun distante, la investigación actual en la estructura de los sistemas tiene la meta de incrementar significativamente el poder computacional aumentando la velocidad de procesamiento, el tamaño de la memoria, la capacidad disponible para la realización de operaciones, una opción factible seria recurrir al paralelismo de los datos, ya que esta ahorraría cantidad significativa de cómputo y de memoria.

Los algoritmos que se utilizan para realizar el Machine Learning (ML), cambian constantemente en la actualidad, siendo en muchas ocasiones más eficientes y eficaces que los anteriores, estos crecen más rápido que la tasa de duplicación de la ley de Moore (la cual dice que cada dos años crece el número de transistores en un microprocesador), sin embargo, estos algoritmos se expresan comúnmente en marcos de alto nivel, lo cual facilita la asignación del programa a arquitecturas innovadoras. Además, existe algo muy importante para la implementación del ML llamado el *TensorFlow*, la cual es una librería o biblioteca de software libre usada para realizar grandes cálculos numéricos a partir de diagramas de flujos de datos de tal forma que se compone de operaciones matemáticas y tensores.

TensorFlow ofrece APIs de alto nivel con el fin de facilitar el desarrollo y preparación de modelos con un excelente rendimiento. Además, contiene un potente compilador que aumenta la velocidad de ejecución de código en cualquier procesador, CPU, GPU o incluso TPU.

También se muestra en el artículo de Gordon Bell y Peter Freeman [1], el cual fue escrito en 1987 (cuando apenas estaba desarrollándose el apogeo de las maquinas en la vida cotidiana), mostrando algunos requerimientos de arquitectura para que la IA sea loa más óptima posible, comunicando que la manera más eficiente a la hora de trabajar la IA es que esta pueda realizar todos sus procesos de manera paralela, además, es necesario contar con por lo menos tres tipos de memorias una principal (Core o circuitos integrados), una secundaria y una terciaria, en las cuales la memoria primaria se encarga de las funciones de la máquina, la memoria secundaria se encargaría del almacenamiento temporal de la información, y la memoria terciaria se encargaría del almacenamiento permanente de la información que esta considere útil, también debe poseer multiprocesadores que funcionarían a modo de "neurona" de la máquina, la siguiente imagen muestra de una manera simplista la estructura de la máquina.

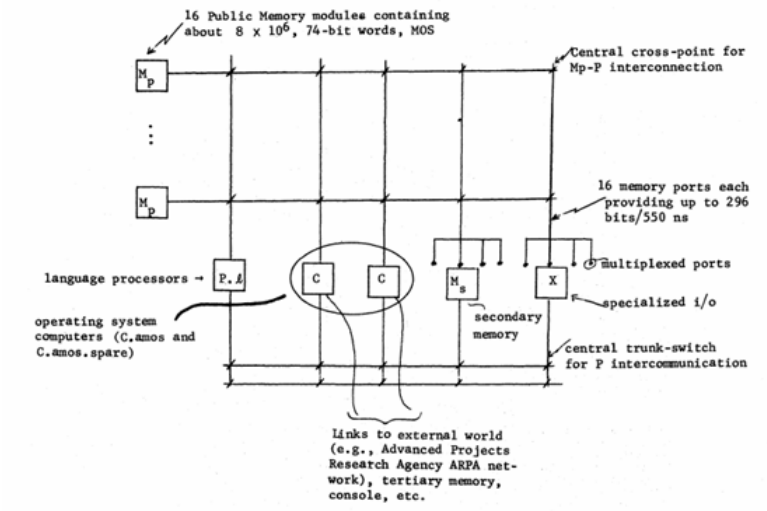

Figure 1. Pueña muestra de la arquitectura propuesta en el articulo.

A continuación, se hablará sobre algunos problemas que trae el Machine Learning directamente a la estructura de los computadores [2].

A. Entrenamiento (Training)

Tanto la capacitación como la inferencia son importantes, pero el trabajo de arquitectura de computadora es insuficiente. la inferencia de datos es más fácil en el entrenamiento. Sin embargo, los requisitos de almacenamiento son mayores en capacitación porque los valores de activación deben ser guardados para su posterior uso en la comparación de datos. No es sorprendente que las primeras incursiones de muchas compañías (incluida la de Google) sobre el hardware de aprendizaje hayan apuntado primero a la inferencia de datos a través de las características.

B. Batch Size

El Batch size permite la reutilización de operandos, sin embargo es muy poco conocido, el conocimiento en el campo del aprendizaje profundo dice que la mejora de la máxima precisión por paso computacional proviene del descenso del gradiente estocástico con impulso a un tamaño de minibatch de 1.
De esta manera, podemos obtener una intuición de descenso de gradiente antes de terminar todo el conjunto de entrenamiento. Hace que el algoritmo sea más rápido y más eficiente. Se llama descenso de gradiente de mini lotes[3].
https://www.coursera.org/lecture/deep-neural-network/mini-batch-gradient-descent-qcogH

C. Sparsity and Embeddings

Sparsity en ML toma muchas formas, que se prestan a enfoques muy diferentes para el soporte de hardware. Los modelos Mixture of Experts (MoE) (El cual cure diferentes regiones de entrada con diferentes tenicas de aprendizaje [4]) entrenan más pesos utilizando menos flops para mayor precisión que los enfoques anteriores, incrustaciones, que transforman grandes espacios dispersos (como vocabularios) en más compactos.

Son representaciones densas adecuadas para operaciones de álgebra lineal, han recibido relativamente poca atención por parte de la comunidad de la arquitectura, pero son clave para aplicaciones textuales como la búsqueda web y traducción.

D. Cuantización y destilación

Se han implementado varias técnicas con gran efecto para proporcionar una inferencia rentable,. Las representaciones aritméticas cuantificadas han demostrado su eficacia en varios aceleradores de inferencia y ahora están disponibles en GPU. Lamentablemente, hay muy poco trabajo en el entrenamiento en aritmética de precisión reducida, y el pequeño trabajo publicado se centra en problemas de una pequeña cantidad de datos como MNIST y CIFAR-10.

E. Networks with Soft Memory

Algunas técnicas de aprendizaje profundo proporcionan una funcionalidad similar a la memoria de acceso aleatorio; hacerlo es se complica preservando la diferenciabilidad. Los ejemplos incluyen Neural Turing Machines, 28–29 redes de memoria, y 30 de atención. La atención le permite a Google Translate mapear desde una palabra en el idioma de destino a la palabra correspondiente o frase en el idioma de origen. Tales mecanismos blandos son caros en comparación con las tradicionales memorias "duras", porque la memoria blanda calcula un promedio ponderado de todas las características de un dato. Una memoria dura simplemente carga una sola característica del dato. No hemos visto investigaciones sobre Implementaciones eficientes o dispersas de estos modelos de memoria blanda.

F. Learning to Learn (L2L)

Los resultados recientes de Zoph et al.32 sugieren que los modelos de redes neuronales pueden buscar por sí mismos y desarrollar mejores modelos de red neuronal. Utilizaron el concurso de reconocimiento de imágenes CIFAR-10 para demostrar que L2L podría igualar la precisión de los mejores modelos desarrollados por expertos de ML.
La búsqueda L2L usó el conjunto de datos CIFAR-10 más pequeño en lugar de ImageNet para hacer que la carga de trabajo computacional sea plausible en el hardware actual, y sin embargo requirió cientos de tarjetas GPU. A pesar del gran crecimiento en el campo, hay una escasez de expertos en tecnologías de Deep Learning. L2L ofrece la posibilidad de que podamos usar muchos más recursos informáticos pero que necesitemos considerablemente menos experiencia humana en el diseño de soluciones. Para aquellos que hacen co-diseño de hardware y software, L2L ofrece la posibilidad adicional de construir mejores sistemas al mismo tiempo, buscando automáticamente nuevas arquitecturas de modelos ML y nuevas arquitecturas de computadoras.

IV. Evolución de la inteligencia artificial en las últimas décadas

El nacimiento de la inteligencia artificial no ha sido oficialmente un hecho memorable o puesto ante el mundo, sin embargo, en el mundo de la computación generalmente se habla que la IA nació con Alan Turing con la creación de máquinas capaces de realizar distintos cálculos ya predefinidos y que pueden usarse en varios campos. Sin embargo, fue gracias a su publicación *Máquinas de computación e inteligencia*, cuando el concepto de inteligencia artificial tomó su fama.

La inteligencia artificial puede verse como una gran herramienta o proceso que expande el uso de las computadoras en su tarea de suplir las tareas del ser humano y es claro que hasta la década de los setenta, el término de inteligencia artificial comenzó a tomar mayor importancia y se hablaba de cómo sería su implementación en el futuro con la ayuda de la mejoría de la capacidad computacional tanto en hardware como en software, cosa que al día de hoy deja de ser una limitación para el desarrollo de la IA, pues muestra cómo se satisface la ley de Moore y con ello el aumento de la capacidad computacional.

Una mirada al futuro según los científicos de la computación de la década de los setenta, para implementar inteligencia artificial requeriría una máquina especializada algo más avanzada que las de esos tiempos. Puede tomarse como referencia la famosa máquina de IMB, la *IMB 360 Model 85*, o máquinas similares y de rendimiento superior que manejaban sólo unos cuantos transistores, las máquinas de hoy en día alcanzan a cumplir con la tal vez ley más incomprendida del mundo de la tecnología, la ley de Moore [5]. Cabe resaltar que como prueba concreta se tienen los últimos procesadores y más comunes en el mercado, como lo son los Intel Core, potentes procesadores que alcanzan una composición de incluso miles de millones de transistores.

Sin embargo, una de las características mencionadas para obtener un computador óptimo para poder trabajar con inteligencia artificial era que el tamaño de memoria de aquel computador ideal, debería trabajar con una memoria primaria de tamaño de aproximadamente $6x10^6$ bits, mientras que al compararlo con una arquitectura de computadores actuales como de 32 bits es claro que puede guardar hasta 2^{3^2} valores, o unos 4,294,967,296 valores. Una arquitectura de 64 bits almacena hasta 2^{6^4} valores o lo que significan 18,446,744,073,709,551,616 valores, una cifra prácticamente imposible de mencionar.

Un computador de la época alcanzaba una memoria de 512 K, al compararlo con una computadora actual que trabaja con una arquitectura comercial de 64 bits, teóricamente alcanza los 7,179,869,184 gigabytes o 16 exabytes, una diferencia incomparable. Esto además de los computadores científicos actuales o incluso los pocos conocidos, pero crecientes computadores cuánticos, que ya se encuentran trabajando de la mano con IA.

Hace unos pocos años, Google le mostró al mundo su nueva innovación en el mundo de la inteligencia artificial, el *TPU* (Tensor Processing Units) o la *Unidad de Procesamiento Tensorial*, chips especializados que viven en el centro de datos de la compañía, las cuales hacen que las tareas realizadas por la IA sean más livianas y eficaces de completar. Estos nuevos chips estan destinados para ser usados en empresas, eso significa tareas como automatizar la calidad de la verificacion de control en las fábricas. Esta tarea difiere en gran escala respecto al uso del hardware que tiene que enviar datos de informacion a través de la internet para dar un análisis del control, trabajar en estos aspectos con machine learning significa una mejoría encuanto a seguridad y tiempo , teniendo unos resultados más rápidos. Vale la pena mencionar que Google no es el único que ha desarrollado aceleradores de ingeligncia artificial, también hacen parte de este avance empresas como ARM, Qualcomm, Mediatek, incluso Nvidia que trabaja con sus GPUs, la cual domina el mercado de entrenamiento de algoritmos.

Los algoritmos que se utilizan para realizar el Maching Learning, cambian constantemente en la actualidad, siendo en muchas ocaciones mas eficientes y eficaces que los anteriores, estos crecen mas rapido que la tasa de duplicacion de la ley de Moore (la cual dice que cada dos años crece el numero de transistores en un microprocesador), sin embargo estos algoritmos se expresan comunmente en marcos de alto nivel, lo cual facilita la asignacion del programa a arquitecturas innovadoras. Además, existe algo muy importante para la implementación del ML llamado el *TensorFlow*, la cual es una librería o biblioteca de software libre usada para realizar grandes cálculos numéricos a partir de diagramas de flujos de datos de tal forma que se compone de operaciones matemáticas y tensores. TensorFlow ofrece APIs de alto nivel con el fin de facilitar el desarrollo y preparación de modelos con un excelente rendimiento. Además contiene un potente compilador que aumenta la velocidad de ejecución de código en cualquier procesador, CPU, GPU o incluso TPU.

Teniendo esta nueva tecnología de procesamiento a la luz del mundo, hasta el momento ha tenido un par de avances o tambien llamadas generaciones. En su primera versión (mostrado en la Figura 2) se enfoncó en la fase de producción de las TPU, debido a la alta demanda sin precedentes de Machine Learning en cuanto a poder computacional se refiere. La segunda generación (mostrado en la Figura 3) se enfonca en una fase de entrenamiento de modelos de Machine Learning de forma ágil.

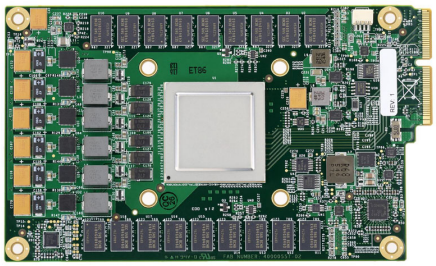

Figure 2. Primera genereación de TPUs de Google, se componía de un sólo chip

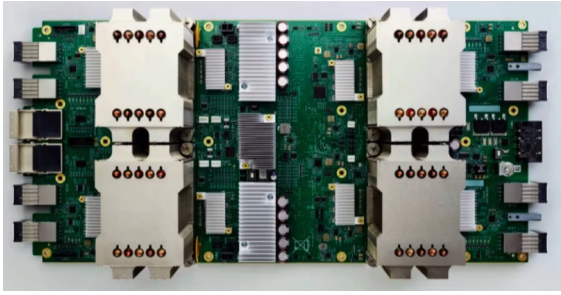

Figure 3. Segunda generación de TPUs, se componía de una placa con cuatro chips.

V. Falacias de hardware ML y escollos

Cuando se utilizan dispositivos con Machine Learning (ML) se encuentran diferentes obstáculos que debemos sortear, y seguir con los problemas mencionados anteriormente. A continuación se podrán encontrar las falacias y escollos, errores, en los que podemos incurrir:

Falacia: Dado el gran tamaño de los problemas de ML, El enfoque del hardware debe ser operaciones por segundo (rendimiento) en lugar de tiempo para la solución (latencia). Al utilizar hardware ML es importante conocer el enfoque que vamos a darle ya que tenemos que decidirnos entre rendimiento o latencia, por tanto, todo dependerá del problema que tenemos por delante, en el caso de problemas de gran tamaño de ML es mejor un enfoque con prioridad en la latencia, siempre respetando los tiempos convergentes y así lograr la precisión deseada.

Falacia: dada una aceleración suficientemente grande, los investigadores de ML Estaría dispuesto a sacrificar un poco de precisión.

Hay que tener cuidado si se decide sacrificar la precisión por el rendimiento, debido que a para los investigadores de ML es necesario crear una aceleración lo suficientemente rápida sin perder precisión, porque tendríamos sistemas que no cumplen los requisitos de lo que necesitamos, por ejemplo, los aplicativos móviles aceptan mayor pérdida de precisión en ganancia de velocidad, caso contrario ocurre en los sistemas de hardware ML. Escollo: Diseñar hardware utilizando modelos recientes.
Recientemente, en los artículos de arquitectura sobre ML se han evaluado las propuestas utilizando MNIST, CIFAR-10 y AlexNet. Sabiendo que MNIST es importante históricamente, con 20 años de uso y además, con uso principalmente para las tareas de programación de los estudiantes. CIFAR-10 fué el concurso de reconocimiento de imágenes que ImageNet reemplazó en 2010 porque se convirtió en demasiado fácil. Incluso AlexNet, que empezó con la revolución CNN. Al ganar ImageNet en 2012, ahora se considera pasado de moda, debido a que solo tiene seis capas de red neuronal, mientras que el campeón de 2017 tiene 150 capas.

Además, tiene la mayor parte de sus parámetros en dos capas totalmente conectadas en la parte superior del modelo, una característica que falta en casi todos los demás modelos de imagen precisos desde AlexNet. Diseñar hardware o evaluar las optimizaciones propuestas basadas en MNIST, CIFAR-10 o AlexNet no es más efectivo o persuasivo que diseñar procesadores de propósito general evaluados con Quicksort, hash, o búsqueda binaria. Además, se desarrollarán suites de referencia para los aceleradores de destitución, similar a la forma en la que los fabricantes de estaciones de trabajo desarrollan la SPEC, hasta que nazcan los puntos de referencia, utilizando los ganadores recientes de ML. Las competiciones como ImageNet aseguran resultados actualizados pero muy cortos de información.

Escollo: Diseñar hardware ML asumiendo que el software ML es intocable. Tomando la ley de Moore y el amplio paralelismo a nivel de instrucción, un arquitecto podía contribuir sin entender lo que el programa está tratando de hacer. Pero existen reglas de suites de referencia como SPEC2006 que no permiten cambios en el código fuente, por lo que estudiarlo no es favorable. Los equipos de hardware que profundizan su trabajo en aprender algunos de los paradigmas y soluciones de la comunidad de ML es más probable que produzca una solución efectiva que una que se adhiera diseño ya creado, no existen requisitos para ejecutar el modelo de la misma forma; Un cálculo más rápido y diferente, pero con la misma precisión es un buen resultado. El objetivo es un sistema completo de hardware y software que resuelva los problemas de una forma más efectiva y mucho más rápida

VI. Machine Learning

Machine Learning aprende por medio de estadísticas, básicamente revisa un puñado de ciertos datos y trata de hallar su significancia, Machine Learning tiene como una de sus principales características el tratar de hallar patrones entre las cosas que significan la misma cosa. Machine Learning puede detectar errores lógicos gracias a suficientes ejemplos para dar con la realización de patrones de reconocimiento de falencias, pero localizarlos podría tratarse incluso de un desafío de ingeniería, pues debería crearse un algoritmo único o especialmente diseñado para tratar ese tipo de cosas, aunque no es algo imposible de realizar. Entonces, en cuanto a dar con una precisa clasificación de errores o fallas, no hay actualmente una manera de explicar el por qué una declaración viene siendo una falla lógica, sin embargo, con una precisa detección se debería ser capaz de explicar el por qué una declaración es una falacia.

Actualmente la cantidad de datos que se obtienen es cada vez mayor, el mundo tiene un desborde de datos (Big data), pero, gracias a esa cantidad de datos podemos tomar una mejor decisión a acontecimientos futuros.

La minería de datos (Data Mining) busca en gran medida encontrar patrones ocultos en las características de los datos, para así ser capaz de predecir futuros comportamientos, pero estos patrones no pueden ser detectados de manera tradicional porque las relaciones son demasiado complejas, o hay una gran cantidad de datos a evaluar [5].

Existen dos tipos de Machine Learning, los cuales son el aprendizaje supervisado y el aprendizaje no supervisado también conocido como Deep Learning. Primeramente, se hablará sobre el aprendizaje supervisado, este es aquel en el cual previamente se le han puesto etiquetas de reconociendo a una serie de características, para que así la maquina sea capaz de observar patrones de estas características, y mediante las etiquetas poder diferenciar los diferentes tipos de información dada. Para ello se utilizan varios métodos de aprendizaje, a continuación, se hablará y explicaran algunos de ellos.

A. Gaussian Naive Bayes:

Para el tratamiento de la inteligencia artificial siempre es importante tener en cuenta a partir de un conjunto de datos verificar a qué tipo de dato pertenece uno entrante, para ello se utiliza la estadística, la cual nos muestra y clasifica el dato entrante al dato más probablemente parecido de nuestro conjunto de datos. Uno de los métodos más sencillos y eficientes a la hora de realizar dicha clasificación es el Gaussian Naive Bayes porque en este método solo se necesita estimar la desviación estándar y estimar la media en los datos de entrenamiento, esta calcula la probabilidad de cada característica realizando una aproximación gaussiana, en la cual no es necesario ajustar los coeficientes mediante procesos de optimización[6].

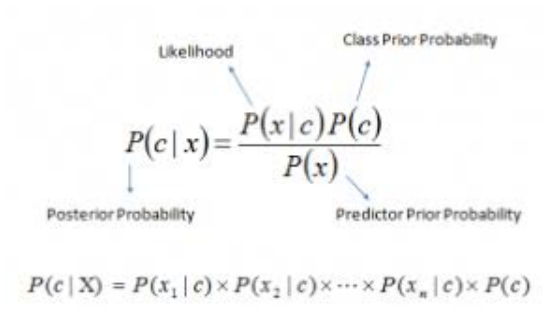

Figure 4. Teorema de Bayes.

Sin embargo, a pesar de ser un clasificador para problemas binarios y problemas múltiples, este funciona de manera más eficiente cuando las características son independientes (no interactúan entre sí), lo cual es muy difícil de encontrar en los conjuntos de datos que se utilizan normalmente, pero, cuando se cumple dicha condición este sería el método más preciso para la clasificación de datos.

B. Decision Trees

Es tipo de aprendizaje es también uno de los más simples, sin embargo, es de los más útiles cuando nos referimos al aprendizaje supervisado, este método consiste esencialmente en realizar una serie de preguntas (generalmente de carácter binario), las cuales se abren las ”ramas” de respuesta, y se volvería a realizar el mismo procedimiento, como se muestra en la imagen No., hasta que se llegue a una clasificación apropiada para el dato entrante [7]. Este método a pesar de ser eficaz,

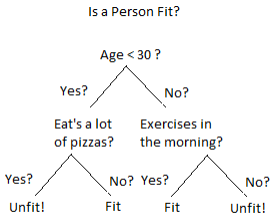

Figure 5. Decision Tree.

no es muy sólido ya que depende en gran medida de la pregunta ”raíz”, porque desde esa pregunta inicial se desprendería las demás ramificaciones, y puede que no se tome en cuenta una o más características de los datos, para ello se implementó la siguiente forma de aprendizaje, el Random Forest

C. Ramndom Forest

Este es uno de los métodos más utilizados gracias a su simplicidad y eficacia, ya que da un gran número de aciertos en la mayoría de las veces, además, que este tipo de aprendizaje sirve para la clasificación y regresión. Este método de aprendizaje utiliza un conjunto de Decision Trees, tomando la pregunta raíz de cada uno de manera aleatoria, dando, así como resultado un conjunto de ”arboles de decisión” con varias cada uno clasificando el dato entrante, así obteniendo una mejor clasificación. Tal y como se muestra en la imagen [8]. A diferencia del Decision Tree, este no tiene el mismo problema, ya que tiene en cuenta

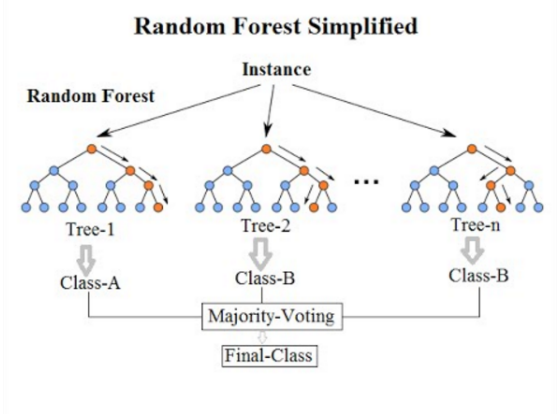

Figure 6. Random Forest.

varias preguntas raíces y no se arraiga a una solo, la única desventaja de este método seria que tiene un orden computacional demasiado alto cuando se tiene una gran cantidad de datos con una gran cantidad de características.

D. Support Vector Machine

Este método de aprendizaje ha sido utilizado por la medicina para el tratamiento de enfermedades, y el descubrimiento de estas antes de que repercutan considerablemente en el cuerpo del paciente, también es utilizado en la actualidad para el reconocimiento facial en los celulares modernos, el reconocimiento de huellas dactilares, etc.
LO que hace que este método sea tan eficaz es el hecho de que este trata de separar los diferentes datos de la manera más óptima, dividiéndola en grupos y tomando solo el grupo en el que coincida estar el dato entrante, esto le hace por medio de kernels, loc cuales pueden ser representados por ecuaciones lineales, o ecuaciones no lineales [9].

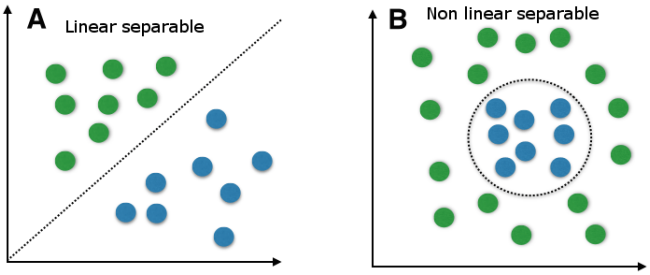

Figure 7. support Vector Machine separando datos de manera lineal y no lineal

El problema que se puede presentar con este método se llama sobre aprendizaje (overlearning), el cual hace que, si se separan los datos de manera demasiado precisa, la inteligencia artificial solo reconocerá unos cuantos tipos de datos, ya que solo aprendió de esos datos, esto bajaría considerablemente la factibilidad de este método.

E. Deep Learning

Este es el método no supervisado, el cual es el más semejante a las neuronas de los mamíferos, las cuales se especializan en detectar determinadas características en las cosas que estos observan, este tipo de aprendizaje es el más reciente, además que representa un acercamiento a lo que es el modo en el que opera el cerebro humano, siendo el área de visión por computador la que más utiliza el Deep Learning. Esta se especializa en encontrar patrones ocultos en las características de los datos, esto le permite ser más eficiente a la hora de clasificar datos, que su contraparte el ML, ya que esta última se centraliza solo en las características entregadas.

CONCLUSIONES

- Se ha cumplido incluso a mayor proporción que la ley de Moore el cumplimiento de las predicciones realizadas por Gordon Bell y Peter Freeman, pero aun así no deja de ser una situación retadora el seguir avanzando en el desarrollo de hardware que satisfaga las crecientes demandas de procesamiento para Machine Learning(ML).
- En el artículo de 1987 se hace la propuesta de todo un ecosistema o ambiente para abordar las labores de ML compuesto por clústeres de equipos de cómputo. Actualmente esto se ve en las ofertas empresariales para ML propuestas por Google Cloud o Amazon Web Services.
- La principal diferencia está dada en la proyección de la capacidad de los procesadores. Bell y Peter hacen su propuesta empleando procesadores de capacidad convencional y dejan la salvedad de que se puede mejorar considerablemente su propuesta de arquitectura si se cuentan con grandes avances. Actualmente los procesadores doblan su capacidad incluso a una tasa superior que la propuesta por "Moore".
- En Colombia se ha adoptado la inteligencia artificial para potenciar las actividades económicas características de nuestra región (Software para predecir la producción de flores y su fecha de cosecha), proveer accesibilidad y mejor calidad de vida a la población en situación de discapacidad (Apps para invidentes que les permite ubicarse geográficamente a través de un asistente con lenguaje natural. También para identificar la denominación de un billete), de igual forma existen importantes aportes desde la academia al cuidado de la salud, por ejemplo, un software para la detección de Parkinson a temprana edad (Investigación realizada por el profesor Fabio de la Universidad Industrial de Santander).
- La consciencia en torno a la relación Arquitectura de Computadores: Alto Desempeño Computacional: Inteligencia Artificial está tomando cada vez más fuerza en Colombia, a través de convenios estratégicos que promueven la puesta en práctica de la IA, investigaciones activas en Universidades sobre supercomputación y espacios para promover el uso, estado del arte e investigación en estos campos (CARLA 2018)

REFERENCES

[1] C. Gordon Bell and P. Freeman, "computer architecture for AI research," December 1972.

[2] J. Dean, D. Patterson and C. Young, "A New Golden Age in Computer Architecture: Empowering the Machine-Learning Revolution," in IEEE Micro, vol. 38, no. 2, pp. 21-29, Mar./Apr. 2018.

[3] James, Vincent, "Google unveils tiny new AI chips for on-device machine learning", 2018; https://www.theverge.com/2018/7/26/17616140/google-edge-tpu-on-device-ai-machine-learning-devkit

[4] "An open source machine learning library for research and production"; https://www.tensorflow.org/?hl=es

[5] Andrew Ng, Kian Katanforoosh, Younes Bensouda Mourri, "Improving Deep Neural Networks": *tuning, Regularization and Optimization*; https://www.coursera.org/lecture/deep-neural-network/mini-batch-gradient-descent-qcogH

[6] Casa Tiempo, "Medellín tiene Centro de Inteligencia Artificial", https://www.eltiempo.com/colombia/medellin/medellin-tiene-centro-de-inteligencia-artificial-212320, Mayo 2018.

[7] A. Altuve, "Empresas tienen en la mira la inteligencia artificial", www.elcolombiano.com, 2019. [Online]. Available: https://www.elcolombiano.com/negocios/empresas/empresas-tienen-en-la-mira-la-inteligencia-artificial-KH9966693.

[8] "¿Qué es la minería de datos?", MuyInteresante.es, 2019. [Online]. Available: https://www.muyinteresante.es/tecnologia/preguntas-respuestas/que-es-la-mineria-de-datos-311477406441.

[9] J. Brownlee, "Naive Bayes for Machine Learning", Machine Learning Mastery, 2016. [Online]. Available: https://machinelearningmastery.com/naive-bayes-for-machine-learning/.

[10] "Decision Trees in Machine Learning – Becoming Human: Artificial Intelligence Magazine", Becoming Human: Artificial Intelligence Magazine [Online]. Available: https://becominghuman.ai/decision-trees-in-machine-learning-f362b296594a.

[11] N. Donges, "The Random Forest Algorithm – Towards Data Science", Towards Data Science, 2018. [Online]. Available: https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd.

[12] Evgeniou, Theodoros and Pontil, Massimiliano. (2001). Support Vector Machines: Theory and Applications. 2049. 249-257. 10.1007/3-540-44673-7-12.