



Una investigación sobre diferentes enfoques de aprendizaje de maquina

Omar García León, Miguel de Nazareth Pineda Becerril, Leonel Gualberto Lopez Salazar, Celina Urrutia Vargas, Flores Pérez Judith Mayte *

RESUMEN

Recientemente el aprendizaje profundo ha obtenido un gran éxito en una variedad de dominios de aplicaciones. Este nuevo campo del aprendizaje automatizado ha crecido rápidamente y ha sido aplicado a la mayoría de los dominios de aplicaciones tradicionales, así como a algunas áreas nuevas que presentan más oportunidades. Se han propuesto diferentes métodos basados en diferentes categorías de aprendizaje, incluido el aprendizaje supervisado, semi-supervisado y no supervisado. Los resultados muestran un rendimiento de vanguardia utilizando el aprendizaje profundo en comparación con los enfoques tradicionales de aprendizaje automático en los campos del procesamiento de imágenes, visión por computadora, reconocimiento de voz, traducción automática, arte, imágenes médicas, procesamiento de información médica, robótica y control, bioinformática, procesamiento del lenguaje natural, ciberseguridad, etc.

ABSTRACT

Deep learning has recently gained great success in a variety of application domains. This new field of machine learning has grown rapidly and has been applied to most of the traditional application domains, as well as some new areas that present more opportunities. Different methods have been proposed based on different categories of learning, including supervised, semi-supervised, and unsupervised learning. The results show cutting-edge performance using deep learning compared to traditional machine learning approaches in the fields of image processing, computer vision, speech recognition, machine translation, art, medical imaging, medical information processing, robotics, and control, bioinformatics, natural language processing, cybersecurity, etc.

Palabras claves: Redes Neuronales Artificiales, Modelos No Paramétricos, Redes Neuronales, Reconocimiento de patrones.

INTRODUCCIÓN

Las redes neuronales artificiales (RNA) son sistemas de procesamiento computacional que están muy inspirados en la forma en que funcionan los sistemas nerviosos biológicos (como el cerebro humano). Las RNA se componen principalmente de una gran cantidad de nodos computacionales interconectados (denominados neuronas), cuyo trabajo se entrelaza de manera distribuida para aprender colectivamente de la entrada a fin de optimizar su salida final.

Las redes neuronales artificiales se aplican en varias situaciones. En particular el planteamiento del problema es descubrir que relación funcional existe entre covariables para lo que se aplica un proceso de entrenamiento iterativo. Su entrenamiento se resume en ensayar iterativamente con perceptrones multicapa en el contexto de análisis de regresión, es decir, para aproximar las relaciones funcionales entre covariables y variables de respuesta se propone buscar la relación entre variables. Por tanto, las redes neuronales se utilizan como extensiones de modelos lineales generalizados. El algoritmo de retropropagación proporciona una opción personalizada de activación y función de error Fig. (1). Teóricamente se puede incluir un número arbitrario de covariables y variables de respuesta, así como de capas ocultas. El documento ofrece una breve introducción a los perceptrones multicapa y la retropropagación resiliente. Los resultados muestran que las redes neuronales se ajustan a los datos durante el proceso de entrenamiento, puede manejar un número arbitrario de variables de entrada y variables de respuesta, no es necesario preespecificar el tipo de relación entre las covariables.

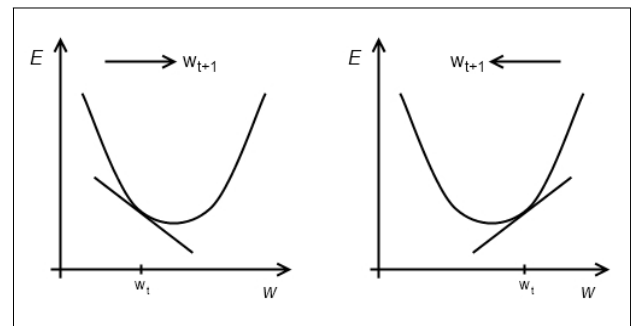


Figura 1.- La idea básica del algoritmo de retropropagación ilustrada para una función de error univariante (Elaboración Propia).

*UNAM, Facultad de Estudios Superiores Cuautitlán, Departamento de Matemáticas, ogarciale@comunidad.unam.mx, Dr. Omar García León, Mtro. Miguel de Nazareth Pineda Becerril, Dr. López Salazar Leonel Gualberto, Dra. Celina Elena Urrutia Vargas, Mtra. Judith Mayte Flores Pérez.



Los dos paradigmas de aprendizaje clave en las tareas de procesamiento de aprendizaje son: el aprendizaje supervisado y no supervisado. El aprendizaje supervisado es el aprendizaje a través de entradas preetiquetadas, que actúan como objetivos. Para cada ejemplo de entrenamiento habrá un conjunto de valores de entrada (vectores) y uno o más valores de salida designados asociados. El objetivo de esta forma de entrenamiento es reducir el error de clasificación general de los modelos, mediante el cálculo correcto del valor de salida del entrenamiento ejemplo.

Una desventaja de la retropropagación es que puede tomar una gran cantidad de iteraciones para converger a la solución deseada. Una alternativa a la retropropagación que se ha utilizado en la clasificación es la red neuronal probabilística (PNN) (Specht (1988), (1990), (1991)), que implica el aprendizaje de un solo paso y se puede implementar directamente en la arquitectura de la red neuronal.

El aprendizaje no supervisado difiere en que el conjunto de entrenamiento no incluye ninguna etiqueta. El éxito suele estar determinado por si la red puede reducir o aumentar una función de costo asociada. Sin embargo, es importante tener en cuenta que la mayoría de las tareas de reconocimiento de patrones generalmente dependen de la clasificación mediante el aprendizaje supervisado.

Existen dos componentes básicos en una red neuronal que son las dos capas básicas de una red: La capa de inputs (primera capa) que consiste de todas las covariantes en neuronas separadas y la capa de outputs (última capa) que consiste de la variable o las variables respuesta (y). Mediante un algoritmo iterativo entre estas dos capas se ajustan los pesos (w_i) que conectan cada capa, de manera que el error entre la variables respuesta (y) y la predicción $\hat{o}(x)$ sea mínimo. Todo este procedimiento de entrenamiento se puede realizar en línea Baluja (1995).

Las redes neuronales convolucionales (RNC) son análogas a las RNA tradicionales en el sentido de que están compuestas por neuronas que se autooptimizan a través del aprendizaje. Cada neurona aún recibirá una entrada y realizará una operación (como un producto escalar seguido de una función no lineal), la base de innumerables RNA. Desde los vectores de imagen sin procesar de entrada hasta la salida final de la puntuación de la clase, toda la red seguirá expresando una única función de puntuación perceptiva (el peso). La última capa contendrá funciones de pérdida asociadas con las clases, y aún se aplican todos los consejos y trucos habituales desarrollados para las RNA tradicionales.

2. DESARROLLO

Tipos de enfoques de DL:

Al igual que el aprendizaje automático, los enfoques de aprendizaje profundo se pueden clasificar de la siguiente manera:

Supervisado, Semi-supervisado o Parcialmente supervisado y No supervisado. Además, existe otra categoría de aprendizaje llamada Aprendizaje por refuerzo (Reinforcement Learning RL) o RL profundo (Deep Reinforcement Learning DRL) que a menudo se discuten bajo el alcance de semi supervisado o, a veces, enfoques de aprendizaje sin supervisión.

Las RNC se componen de tres tipos de capas. Estas son capas convolucionales, capas de agrupación y capas totalmente conectadas. Cuando estas capas se apilan, se ha formado una arquitectura RNC. En la Fig. 2 (O'Shea 2015) se ilustra una arquitectura CNN simplificada.

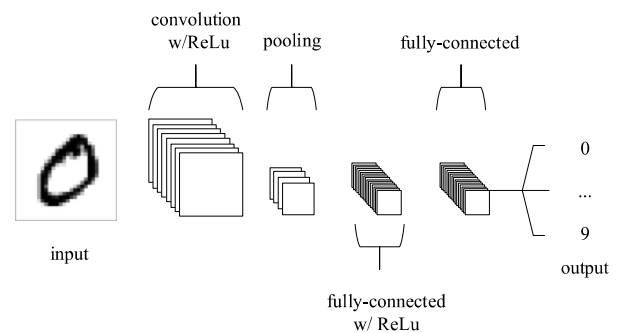


Figura 2.- Ilustra la arquitectura CNN simplificada (O'Shea 2015).

El aprendizaje supervisado es una técnica de aprendizaje que utiliza etiquetas de datos. En el caso de enfoques DL supervisados, el entorno tiene un conjunto de entradas y salidas correspondientes $(x_t, y_t) \sim p$. Por ejemplo, si para la entrada x_t , el agente inteligente predice $\hat{y} = t f(x)$, el agente recibirá un valor de pérdida $l(y, \hat{y})$.

Los sistemas de aprendizaje no supervisados son aquellos que pueden hacerlo sin la presencia de etiquetas de datos. En este caso, el agente aprende la representación interna o características importantes para descubrir relaciones o estructuras desconocidas dentro de los datos de entrada. A menudo, la agrupación, la reducción de la dimensionalidad y las técnicas generativas se consideran enfoques de aprendizaje no supervisados. Hay varios miembros de la familia del aprendizaje profundo que son buenos para la agrupación en clústeres y la reducción de dimensionalidad no lineal, incluidos Codificadores automáticos (AE), máquinas Boltzmann restringidas (RBM) y GAN de reciente desarrollo. Además, los RNN, como LSTM y RL, también



se utilizan para el aprendizaje no supervisado en muchos dominios de aplicación Arulkumaran(2017).

A través de este método simple de transformación, las CNN pueden transformar la entrada original capa por capa utilizando técnicas convolucionales y de reducción de muestreo para producir puntajes de clase con fines de clasificación y regresión.

3. RESULTADOS Y ANÁLISIS

3.1 Descenso de gradiente

El PMC calcula la ecuación (1):

$$o(x) = f\left(w_0 + \sum_{i=1}^n w_i x_i\right) = f(w_0 + \mathbf{w}^T \mathbf{x}) \quad (1)$$

$o(\mathbf{x})$ = Pronóstico
 w_0 = Intersección
 $\mathbf{w}=(w_1, \dots, w_n)$
 $\mathbf{x}=(x_1, \dots, x_n)$

donde w_0 denota la intersección, $\mathbf{w}=(w_1, \dots, w_n)$ el vector que consta de todos los pesos sinápticos sin la intersección y $\mathbf{x}=(x_1, \dots, x_n)$ el vector de todas las covariables. Esta función es matemáticamente equivalente a la de MLG (Modelos Lineales Generalizados) con función de enlace f^{-1} . Por lo tanto, todos los pesos (w_0, w_1) calculados en este caso son equivalentes a los parámetros de regresión del MLG.

Para aumentar la flexibilidad del modelado, se pueden incluir capas ocultas. Sin embargo, Hornik et al. (1989) demostró que una capa oculta es suficiente para modelar cualquier función continua a intervalos. Tal PMC con una capa oculta que consta de J neuronas ocultas calcula la ecuación (2):

$$\begin{aligned} o(\mathbf{x}) &= f\left(w_0 + \sum_{j=1}^J w_j \cdot f\left(w_{0j} + \sum_{i=1}^R w_{ij} x_i\right)\right) \\ &= f\left(w_0 + \sum_{j=1}^J w_j \cdot f\left(w_{0j} + \mathbf{w}_j^T \mathbf{x}\right)\right), \quad (2) \end{aligned}$$

$o(\mathbf{x})$ = Predicción
 w_0 = Intersección
 w_{0j} = Intersección en neuronal j^{th}
 $\mathbf{w}_j=(w_{1j}, \dots, w_{nj})$
 $\mathbf{x}=(x_1, \dots, x_n)$

donde w_0 denota la intersección de la neurona de salida y w_{0j} la intersección de la neurona oculta. Además w_j , denota el peso sináptico correspondiente a la sinapsis que comienza en la j^{th} neurona oculta.

El enfoque de descenso de gradiente es un algoritmo de optimización de primer orden que se utiliza para encontrar los mínimos locales de una función objetivo. Esto se ha utilizado para entrenar a las RNA en las últimas dos décadas con éxito. El algoritmo I explica el concepto de descenso de gradiente:

Algoritmo I. Gradiente de descenso

Entradas: función de pérdida ε , tasa de aprendizaje η , conjunto de datos X , y y el modelo $F(\theta, x)$
Salidas: Óptimo θ que minimiza ε
REPETIR hasta converger:

$\tilde{y} = F(\theta, x)$
 $\theta = \theta - \eta \cdot$

FIN

3.2. Descenso de gradiente estocástico (SGD)

Dado que un tiempo de entrenamiento prolongado es el principal inconveniente del enfoque tradicional de descenso en gradiente, el enfoque SGD se utiliza para entrenar redes neuronales profundas (DNN) (Bottou 2012). Algoritmo II explica SGD en detalle.



REFERENCIAS Y BIBLIOGRAFÍA

Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). A brief survey of deep reinforcement learning. arXiv preprint arXiv:1708.05866.

Baluja, S., & Pomerleau, D. (1995). Encouraging distributed input reliance in spatially constrained artificial neural networks: Applications to visual scene analysis and control, 20(1), 23-38.

Bottou, L. (2012). Stochastic gradient descent tricks. In Neural networks: Tricks of the trade (pp. 421-436). Springer, Berlin, Heidelberg.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. Neural networks, 2(5), 359-366.

O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458.

Specht, D. F. (1988, July). Probabilistic neural networks for classification, mapping, or associative memory. In ICNN (pp. 525-532).

Specht, D. F. (1990). Probabilistic neural networks. Neural networks, 3(1), 109-118

Specht, D. F. (1991). A general regression neural network. IEEE transactions on neural networks, 2(6), 568-576

INFORMACIÓN ACADÉMICA

Omar García León: Licenciado en Administración Financiera por el Instituto Tecnológico y de Estudios Superiores de Monterrey. Maestría en Empresa en la Universidad de Barcelona, Doctor en la Universidad de Barcelona.

Miguel De Nazareth Pineda Becerril: Maestría en Ingeniería, Orientación: Mecánica, Campo Disciplinario: Metalmeccánica, Universidad Nacional Autónoma de México, Facultad de Estudios Superiores Cuautitlán Carrera Ingeniería Mecánica Eléctrica (Industrial), Universidad Nacional Autónoma de México, Facultad de Estudios Superiores Cuautitlán.

López Salazar Leonel Gualberto: Licenciado en Matemáticas Aplicadas y Computación por la FES Acatlán, Maestro en Gestión de Tecnologías de Información Universidad TEC MILENIO y doctorado en educación por parte de la UCI. Tutor de cursos y diplomados en línea para la DGTIC en el programa de capacitación de profesores en "Herramientas Tecnológicas para la enseñanza".

Celina Elena Urrutia Vargas: Ingeniera Química Industrial, egresada del Instituto Politécnico Nacional. Maestría y Especialidad en Educación, Doctorado en Educación. Ponente en el Diplomado "Enseñanza de las Matemáticas con el uso de las Nuevas Tecnologías" módulo I. "Reconocimiento Sor Juana Inés de la Cruz año 2020".

Judith Mayte Flores Pérez: Mae UPIITA. Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas. Ingeniera en Mecatrónica. SEPI ESE. Sección de Estudios de Posgrado e Investigación. Escuela Superior de Economía. Maestría en Ciencias Económicas, con especialidad en Desarrollo Económico. Estancia de investigación en la Universidad de Santiago de Compostela.

