

¿Cómo entienden las computadoras nuestros sentimientos?

How do computers understand our feelings?

Por Kevin Mejía González, Griselda Areli Matías Mendoza, Yulia Ledeneva

Resumen: Expresar emociones o sentimientos en las redes sociales es una actividad cotidiana; sin embargo, ¿una computadora puede entender lo que sentimos? En este artículo explicamos matemáticamente cómo una computadora procesa cientos de opiniones en cuestión de segundos y determina si una opinión cuenta con alguna emoción y su polaridad o valoración como positiva, negativa o neutra.

Palabras clave: opiniones, aprendizaje supervisado, aprendizaje no supervisado, transformadores, procesamiento del lenguaje natural, emociones, polaridades.

Abstract: Expressing emotions or feelings on social media is an everyday activity; however, can a computer understand how we feel? In this article, we explain mathematically how a computer processes hundreds of opinions in a matter of seconds and determines whether an opinion contains any emotion and its polarity or sentiment—positive, negative, or neutral.

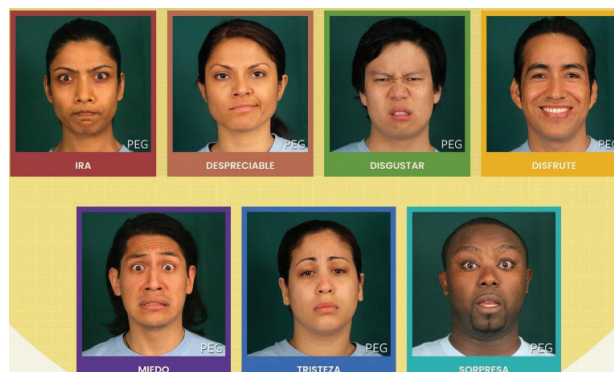
Keywords: opinions, supervised learning, unsupervised learning, transformers, natural language processing, emotions and polarities.

Los sentimientos, innatos en la humanidad, nos ayudan a procesar diversas situaciones de nuestra vida diaria, agradables o no. Con la creación del internet, específicamente la versión 2.0, estos sentimientos no solo quedan en el subconsciente, sino que han podido ser compartidos en las redes sociales.

Esto ha influido en el crecimiento exponencial de opiniones que se difunden en la web y, por ello, ha surgido la necesidad de interpretar esos mensajes. Hoy existen diversas investigaciones para saber cómo es que las computadoras reconocen el lenguaje de los humanos y una de estas aproximaciones es el análisis de sentimientos. Liu (2012) lo describe como el estudio para conocer,

de forma automática, la carga sentimental (Figura 1), también llamadas polaridades positivas, negativas o neutrales.

Figura 1. Las 7 emociones básicas de Paul Ekman



Fuente: Paul Ekman Group (s/f).



Tecnología

Cuando nos referimos a un reconocimiento automático, inferimos que las computadoras “entienden” nuestras emociones de forma textual mediante procesos matemáticos y semánticos. Por esa razón existen diversos métodos por los que la tecnología puede decirnos si una opinión tiene una carga emocional o no.

Uno de los métodos más comunes es el uso de lexicones (diccionarios), cuya estructura es similar a nuestro lenguaje, es decir, contienen sustantivos, adjetivos, verbos, pronombres, adverbios, preposiciones y conjunciones. Para generar un análisis de sentimientos se requiere conformar la sintaxis de cada idioma, debido a que estos diccionarios asignan una valencia a cada palabra a partir de la composición de las frases (Figura 2). Esto ayuda a identificar si un comentario tiene polaridad positiva, negativa o neutral.

Figura 2. Ejemplo de valencias aplicadas a una opinión



Fuente: Ilustración propia.

Las valencias que se extraen desde un diccionario indican el nivel de polaridad de cada palabra w_i que contiene una opinión. Para ello, hemos formulado una ecuación que consiste en una sumatoria básica, donde i es el valor mínimo de palabras, n el límite y w el peso sináptico o la frecuencia ($w_1 + w_2 + w_3 + \dots$) (Figura 3).

Figura 3: Ecuación para determinar la polaridad de una oración.

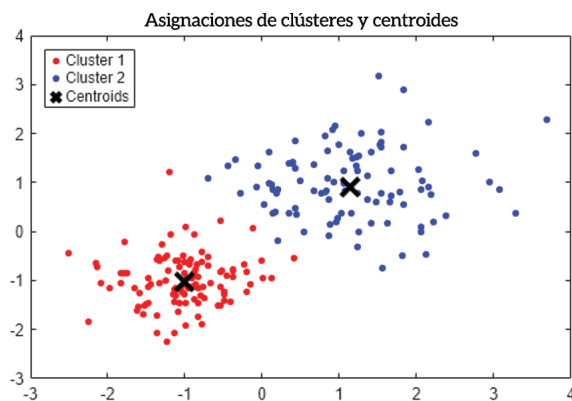
$$\sum_{i=1}^n w_i$$

Por otro lado está la ambigüedad de significados. Si ponemos atención a la frase “La niña se sienta en el banco”, *sienta* y *banco* pueden hacer alusión a un mueble o a una institución financiera en el caso de la primera; la segunda puede corresponder al verbo sentir y a tomar asiento.

Para afrontar esta problemática, se aplican técnicas de *machine learning* más sofisticadas: modelos no supervisados y supervisados. Los primeros tienen por objetivo encontrar mediante operaciones matemáticas un conjunto de patrones dentro de los datos, con el fin de agrupar esos datos en etiquetas.¹ Si tomamos como ejemplo los comentarios en redes sociales, algunos tienen una carga de felicidad, agrado o entusiasmo (positivas) o enojo, frustración u odio (negativas). Como humanos, entendemos esas expresiones, pero los modelos de aprendizaje no supervisado determinan el sentimiento con ayuda de agrupamientos en *k*-medias (kmeans), DBSCAN (agrupamiento espacial basado en densidad para aplicaciones con ruido) y LDA (asignación de Dirichlet latente), entre otras tecnologías predictivas.

Con esta herramienta es posible observar los centroides (Figura 3) que, debido a la naturaleza del algoritmo, fueron seleccionados de manera aleatoria. Por ello, a partir de una ecuación que mide la similitud entre los distintos datos dispersos en la gráfica, el algoritmo determina a qué centroeide (o etiqueta) pertenece cada dato.

Figura 3. Ejemplo de dos centroides escogidos aleatoriamente en un modelo kmeans



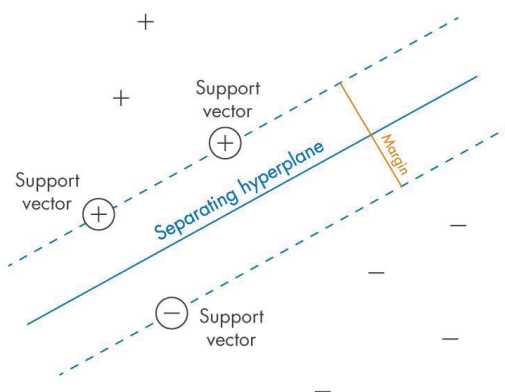
Fuente: Centro de ayuda de Matlab.

¹Las etiquetas sirven para conocer a qué conjunto pertenece un subconjunto de datos. Si ponemos de ejemplo las opiniones, se puede decir que ese conjunto de datos pueden ser un subconjunto de positivos, negativos o neutrales.

A su vez, los modelos supervisados sí requieren un conjunto de datos previamente etiquetados, ya sea por humanos o por algoritmos no supervisados para establecer patrones inmersos en la información de entrenamiento. Posteriormente, se evalúan con un conjunto de datos reales no etiquetados pero similares a los patrones de la prueba, de manera que estos algoritmos deducen a qué etiqueta pertenecen los nuevos datos. Hagamos una analogía, cuando nos encontramos en una situación difícil, como aprobar un examen que vale la calificación total de todo el curso, nos entrenamos mentalmente con ejemplos parecidos a los que pueden venir en la prueba. Una vez terminado ese examen, un experto (profesor) evalúa la cantidad de aciertos o errores al contestar y, con base en eso, asigna una calificación. Pasa lo mismo con un algoritmo supervisado; al final, nosotros evaluamos la fiabilidad de nuestro método.

Existen diversos algoritmos supervisados, como la regresión logística (LR), las máquinas de soporte vectorial (SVM), los árboles de decisión (DT), el bosque aleatorio (RF), los k vecinos más cercanos (KNN), *naïve bayes* (NB) y las redes neuronales (NR). Aquí un ejemplo: tenemos un conjunto de datos dividido por una línea llamada hiperplano de decisión, la cual indica qué tanto deben estar separados los datos de un grupo con respecto a los otros (Figura 4).

Figura 4. Representación gráfica de la funcionalidad de una svm

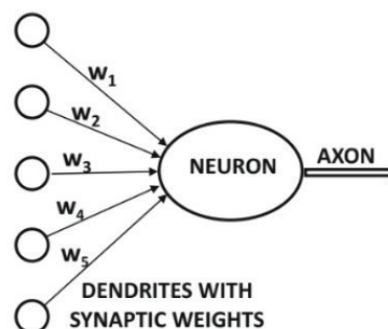


Fuente: Introducción a support vector machines (svm).

Los desafíos en el análisis de sentimientos han evolucionado con el tiempo y los algoritmos supervisados lo han hecho de igual forma. Con esta evolución ha llegado una disciplina llamada aprendizaje profundo, que marca el renacimiento y expansión de lo que conocemos como inteligencia artificial.

El aprendizaje profundo utiliza redes neuronales artificiales con múltiples capas para aprender representaciones jerárquicas y de los datos sin depender de los humanos. Una red neuronal (RN), según Aggarwal (2018), es un modelo computacional inspirado en la estructura y funcionamiento del cerebro humano, compuesto por unidades de procesamiento simples llamadas neuronas artificiales, organizadas en capas y conectadas mediante pesos sinápticos. Tales redes son capaces de aprender vínculos complejos entre entradas y salidas a través de un proceso de ajuste iterativo de pesos (Figura 5).

Figura 5. Representación básica de una red neuronal



Fuente: Aggarwal, 2018.



Estos pesos (w) y entradas convergen en lo que se conoce como perceptrón (neurona artificial), cuyo umbral actúa como decisión cuando la operación se realiza entre los pesos y las entradas es válida para ser una salida.

En el aprendizaje profundo, estas redes neuronales se encuentran conectadas con otras en cantidades muy grandes, formando capas internas de redes en donde las salidas de unas son las entradas de otras.

Además, existen diversos algoritmos de aprendizaje supervisado, como las redes neuronales profundas (DNN), redes convolucionales (CNN), redes recurrentes (RNN), transformadores e inteligencias artificiales generativas, como BERT, Open Roberta Lab, Chat-GPT, Gemini, Deep-Seek, etcétera. La ventaja de estos radica en su capacidad de adaptarse para desempeñar tareas específicas. Los transformadores –que también constituyen un tipo de red neuronal– han sido modificados para una amplia variedad de aplicaciones, que abarcan desde la visión por computadora, el procesamiento de audio y voz, hasta campos como la biología, la química, las finanzas y el análisis de sentimientos. En el análisis de sentimientos, los modelos basados en transformadores han demostrado un desempeño superior al de los algoritmos tradicionales, tanto supervisados como no supervisados, debido a su capacidad de comprender de manera más precisa el contexto presente en una opinión.

LAS COMPUTADORAS “ENTIENDEN” NUESTRAS EMOCIONES DE FORMA TEXTUAL MEDIANTE PROCESOS MATEMÁTICOS Y SEMÁNTICOS; ASÍ, LA TECNOLOGÍA PUEDE DECIRNOS DE FORMA AUTOMÁTICA SI UNA OPINIÓN TIENE UNA CARGA EMOCIONAL O NO

Regresando un poco al ejemplo de banco y sienta, los transformadores son capaces de reconocer las palabras anteriores o posteriores de un término que pueda tener varios significados. La ambigüedad se resuelve con el mecanismo de atención, el cual va generando su conocimiento a partir de otros ejemplos que se pueden llegar a ver en un conjunto de datos de preentrenamiento.² Sin embargo, este tipo de redes neuronales no analizan únicamente los datos en la fase previa, sino que utiliza un conjunto de datos de entrenamiento. Este proceso sirve para que un transformador se afine, obligando al modelo a que, ade-

² Este preentrenamiento es el punto clave de un transformador, pues se ve obligado a aprender con respecto a un lenguaje en general orientado a un contexto en común, como el análisis de sentimientos.

más de lo aprendido, se ajuste a una tarea específica, como analizar emociones en un conjunto de opiniones y, a su vez, reconocer una polaridad específica, como un ataque de odio.

Una vez entendido el trasfondo de los algoritmos que se han desarrollado hasta nuestros días, queda por responder a la pregunta inicial: ¿cómo entienden nuestros sentimientos las computadoras? Si recapitulamos un poco, cuando escribimos una opinión, no solo podemos copiarla y pegarla en un modelo para que nos diga qué significa esa idea en el ámbito sentimental, sino que debemos hacer comprender a una computadora el texto de forma numérica, pues todos los algoritmos anteriores se basan en fórmulas matemáticas.

Para esto, existen diversos métodos que abstraen lo textual a lo numérico.

1. **Bolsa de palabras.** Representa un texto en relación con las palabras que contiene. Únicamente se coloca un valor de 1 si se encuentra la palabra y un valor de 0 en caso contrario.
2. **TF-IDF** (frecuencia de términos–frecuencias de términos inversa). Comencemos con TF (*term frequency*), un método que permite identificar la relevancia de una palabra dentro de un documento en comparación con el conjunto total de opiniones. Por su parte, IDF (*inverse document frequency*) mide qué tan poco común o rara es una palabra dentro

EL ANÁLISIS DE SENTIMIENTOS HA EVOLUCIONADO Y LOS ALGORITMOS SUPERVISADOS LO HAN HECHO DE IGUAL FORMA; HOY, EL APRENDIZAJE PROFUNDO SE CONOCE COMO INTELIGENCIA ARTIFICIAL

de todo el corpus. Para entender de mejor manera, el TF divide el número de veces que un término t aparece en d (opiniones) y eso entre el total de palabras que tenemos en d . Ahora, el IDF es el cálculo de log del total de opiniones N entre el número de veces que t aparece en N (el exponente para llegar al resultado). Una vez que tenemos estos valores, los multiplicamos. Este método, por tanto, nos indica que entre más rara es una palabra más relevante es para todo el conjunto de opiniones; tiene mayor peso TF-IDF.

3. **N-gramas.** Es una secuencia continua de N elementos –que pueden ser palabras, caracteres o sílabas– extraída de un texto. Estos se representan como vectores segmentados según el valor de N , lo cual permite analizar las relaciones entre los elementos dentro de la secuencia. En la frase: “No me gusta el servicio del banco”, si el valor de N es 1, dividimos la opinión en

la unidad más básica de texto que es un *token* [No, me, gusta, el, servicio, del, banco]. Si $N=2$ gramas [No me, me gusta, gusta el, el servicio, servicio del, del banco]. Esta técnica nos ayuda a conocer el contexto de un comentario o la desambiguación de este.

4. **Word embeddings (incrustaciones de palabras).** Un *embedding* es una representación numérica de las palabras que codifica el significado de estas, así como los n -gramas, y estas deben mostrarse en un vector. Esta técnica es útil para identificar el contexto dentro del conjunto. En los casos de “El banco da préstamos” y “Voy al cajero del banco del centro”, banco tiene ya un significado más específico, por la frecuencia del contexto en el que se emite. Si un término se repite n veces dentro de un conjunto de k opiniones similares, entonces dicho término pierde por completo el sentido original, dificultando que un modelo de aprendizaje supervisado pueda interpretarlo correctamente. *Word2Vec*, *GloVe* y *FastText* son algunos ejemplos de estas técnicas.

5. **Context embeddings (incrustaciones de contexto).** Esta técnica es una evolución de las *word embeddings*. Ahora, en lugar de solo tomar las palabras de un n término por opinión con respecto a sus palabras cercanas, se considera todo el conjunto. Esto ayuda a conocer a la desambiguación, pues la base está en los transformadores.

Si tomamos en cuenta que los conjuntos de opiniones tienen una cantidad mayor de ejemplos de los mencionados, esas representaciones deben plasmarse en matrices y, ahora sí, una computadora podría descifrarlas para obtener las emociones y las polaridades. Sin embargo, aún debemos revisar las etiquetas, pues sin ellas los modelos no podrían saber a qué se refiere una opinión cuando se encuentran en la fase de entrenamiento (Tablas 1 y 2).

Tabla 1. Ejemplo de opiniones con su etiqueta P = positiva, N=negativa y NEU=neutral

Opinión	Etiqueta
Me encanta el servicio del banco.	P
El cielo es azul.	NEU
Odio el puesto de tacos, no tiene bancos donde sentarse.	N
De los artículos de investigación aprendo mucho. ¡Qué emoción!	P



Tecnología

Tabla 2. Representación de una matriz basada en bolsa de palabras

encantan	servicio	banco	cielo	azul	odio	puesto	tacos	sentarse	artículos	investigación	aprendo	emoción	etiqueta
1	1	1	0	0	0	0	0	0	0	0	0	0	P
0	0	0	1	1	0	0	0	0	0	0	0	0	NEU
0	0	1	0	0	1	1	1	1	0	0	0	0	N
0	0	0	0	0	0	0	0	0	1	1	1	1	P

En los cuatro ejemplos de la tabla 1, las opiniones etiquetadas con respecto a su polaridad. Las palabras *el, de, del, es, no, los y que* no proporcionan información relevante para el significado completo; se les conoce como stopwords (palabras vacías) y es común que sean eliminadas en tareas como el análisis de sentimientos.

En la tabla 2, se muestra la bolsa de palabras de las opiniones mencionadas en la tabla 1. Se coloca el valor 1 si en el conjunto total se encuentra el subconjunto de palabras de cada opinión, en caso contrario el valor es de 0.

Los vectores generados en un embedding simplificado permiten representar cada palabra como un punto en un espacio multidimensional. Esto explica por qué términos como *encantan* y *emoción* aparecen con valores muy similares: su cercanía en ese espacio revela su semejanza semántica, tal como ocurre en algoritmos como *kmeans*, donde los puntos más próximos tienden a agruparse bajo la misma etiqueta (Tabla 3).

Tabla 3. Representación simplificada de las opiniones en word embeddings

Palabra	Vector de embedding
encantan	[0.72, 0.31, 0.60]
servicio	[0.48, 0.20, 0.44]
banco	[0.45, 0.10, 0.40]
cielo	[0.22, 0.55, 0.71]
azul	[0.30, 0.65, 0.80]
odio	[-0.80, 0.25, -0.10]
tacos	[-0.20, 0.15, 0.10]
aprendo	[0.62, 0.40, 0.70]
emoción	[0.70, 0.55, 0.68]

Esta idea matemática contrasta con la manera en que los humanos procesamos el lenguaje, pues nuestro cerebro aprende a interpretarlo

desde que escuchamos por primera vez una palabra, internalizando patrones, contextos y emociones que una máquina no experimenta de forma auténtica. Aunque la ia solo ejecuta operaciones matemáticas sin la capacidad de “sentir”, su acelerado desarrollo sugiere que quizá no estemos tan lejos de modelar computacionalmente aspectos cada vez más profundos de la experiencia humana y emocional.

Si en algún momento logramos aproximar, aunque sea parcialmente, la capacidad humana de interpretar y responder a emociones en un sistema artificial, podríamos considerar que hemos alcanzado un hito significativo en el desarrollo de estas tecnologías.

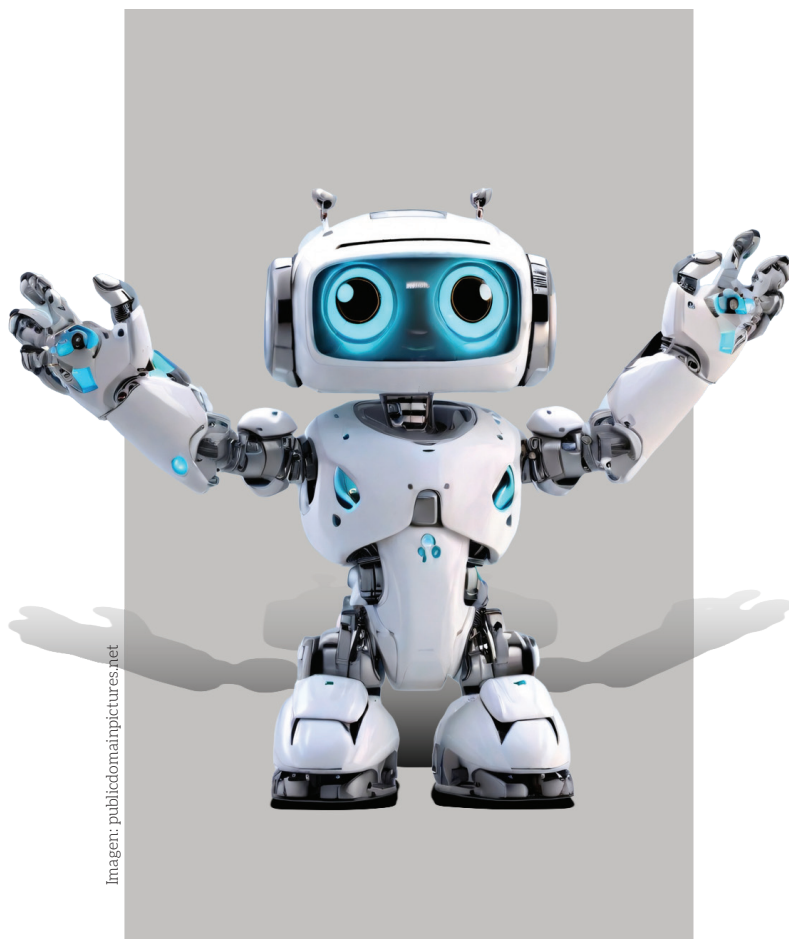


Imagen: publicdomainpictures.net

Referencias

- Aggarwal, Charu C. (2018). *Neural networks and deep learning*. Springer International. <https://www.google.com.mx/books/edition/Neural_Networks_and_Deep_Learning/achqDwAAQBAJ?hl=es-419&gbpv=1&dq=neural+networks+books&printsec=frontcover>.
- Centro de ayuda de Matlab. *Agrupamiento de kmeans*. <<https://la.mathworks.com/help/stats/kmeans.html>>.
- Introducción a *support vector machines (svm)* (s/f).. <<https://la.mathworks.com/discovery/support-vector-machine.html>>.
- Liu, Bing (2012). "Sentiment analysis: a fascinating problem", en *Sentiment analysis and opinion mining*. <https://link.springer.com/capter/10.1007/978-3-031-02145-9_1>.
- Paul Ekman Group (s/f). *What are emotions?* <<https://www.paulekman.com/universal-emotions/>>.



Kevin Mejía González es Maestro en Ciencias de la Computación por la UAEMéx. Estudia el Doctorado en Ciencias de la Computación. Sus temas de interés son inteligencia artificial, procesamiento del lenguaje natural, computación evolutiva y análisis de sentimientos en redes sociales.



Griselda Areli Matías Mendoza es Doctora en Ciencias de la Computación por la UAEMéx. Actualmente es profesora de asignatura y posgrado en el Centro Universitario UAEMéx Tianguistenco, miembro del SNII, nivel I. Sus líneas de investigación son inteligencia artificial, procesamiento del lenguaje natural, computación evolutiva y tecnologías del lenguaje aplicadas a lenguas indígenas.



Yulia Ledeneva es Doctora en Ciencias de la Computación por el CIC-IPN. Actualmente, es profesora-investigadora de la UAEMéx, miembro del SNII nivel II y autora de más de 85 publicaciones científicas en revistas de alto impacto y congresos internacionales. Sus áreas de investigación son inteligencia artificial, procesamiento del lenguaje natural, algoritmos genéticos y métodos basados en grafos.