



TIES Revista de Tecnología e Innovación en Educación Superior

ANÁLISIS DE REDES SOCIALES EN TORNO A LA EMERGENCIA SANITARIA DE COVID-19

<https://doi.org/10.22201/dgtic.26832968e.2022.6.1>

Everardo Bárcenas Patiño, Universidad de
Grenoble

ORCID: 0000-0002-1523-1579

Guillermo Gilberto Molero Castillo, Universidad de
Guadalajara

ORCID: 0000-0002-6330-6408

Agustín Daniel Ambrosio Aguilar, Facultad de
Ingeniería, UNAM

Rocío Aldeco Pérez, Universidad de Southampton,
Reino Unido, ORCID:

ORCID: 0000-0002-7003-2724

<https://www.ties.unam.mx/>

Fecha de recepción: noviembre 23, 2021 • Fecha de publicación: marzo, 2023

Marzo 2023 | número de revista 7 • ISSN 2683-2968

Dirección de Sistemas y Servicios Institucionales, Dirección General de Cómputo y de
Tecnologías de Información y Comunicación, UNAM

Esta obra está bajo licencia de Creative Commons
Atribución-No Comercial 4.0 Internacional (CC BY-NC 4.0)

Análisis de redes sociales en torno a la emergencia sanitaria de COVID-19

Social media Analysis of COVID-19 health emergency

Resumen

El amplio incremento en el uso de redes sociales, también conocidas como medios sociales, ha detonado estudios de la información publicados a través de las redes sociales con respecto a diversos dominios: elecciones para cargos públicos, desastres naturales, guerras, estados de salud, por mencionar algunos. En este artículo se describe el estado del arte sobre el análisis de redes sociales. Así mismo, se propone un estudio para la identificación de patrones de ubicación de publicaciones en redes sociales con respecto al nivel de contagios *COVID-19* en determinadas zonas de la Ciudad de México. Estos patrones pueden ser de utilidad para los actores políticos, sanitarios y/o gubernamentales con la finalidad de tomar medidas de prevención y/o emergencia sanitaria.

Palabras clave:

Redes Sociales, Detección de Patrones, *COVID-19*, Inteligencia Artificial, Aprendizaje Computacional

Abstract

The wide and fast adoption of social media technology has triggered recent studies on social media patterns in several domains: poll elections, natural hazards, wars, health conditions, just to mention some. In this paper, we describe the current state-of-the-art of social media pattern detection. Furthermore, we also propose a method to associate social media post locations with *COVID-19* infection rate in particular Mexico City areas. These patterns may be useful for policy makers, health and government authorities to prevent and assist during *COVID-19* contingency.

Keywords:

Social Media, Pattern Detection, *COVID-19*, Artificial Intelligence, Machine Learning

Análisis de redes sociales en torno a la emergencia sanitaria de COVID-19

Introducción

Las redes sociales, también conocidas como medios sociales, son plataformas digitales que permiten a los usuarios, compartir una relación, un interés común o intercambiar información entre sí. Entre la amplia variedad de redes sociales aparecen: *Facebook*, *Twitter*, *Instagram*. Con el rápido incremento del acceso a equipos de cómputo con internet, la cantidad de usuarios ha crecido enormemente, y se espera que se acelere aún más. En 2020, se estimaron 3,600 millones de usuarios

de redes sociales en el mundo [1]. Esta cantidad de usuarios representa aproximadamente el 46% de la población mundial [2]. Más aún, se proyectan 4,410 millones de usuarios para 2025. En el caso de México, en 2020 se estimaron 80.8 millones de usuarios, el 64% de su población, grosso modo [3]. Se estima un incremento para el año 2025 que resulte en 95.22 millones de usuarios en México. En la Tabla 1, se listan los usuarios de redes sociales en algunos países de lengua castellana.

Tabla 1. Statista, “Number of social network users worldwide 2017 to 2025”, 2021.
[Online]. Available: <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>. [Consultado en abril 27, 2022].

País	Usuarios	Población	Porcentaje poblacional
Chile	16	19.12	83.68%
Perú	27	32.97	81.89%
Colombia	39	50.88	76.65%
Argentina	34	45.38	74.92%
México	80.88	126	64.19%
España	29	47.39	61.19%

La información producida por los usuarios de redes sociales ha demostrado ser de gran utilidad en diversos contextos. Por ejemplo, en la predicción de resultados electorales [4], identificación de daños en desastres naturales [5], detección de estados de salud mental, tales como depresión [6], entre otros. En este artículo se propone un método basado en aprendizaje computacional para detectar la ubicación de publicaciones en redes sociales. En particular, las relacionadas con la contingencia sanitaria causada por COVID-19. La ubicación de estas, se puede vincular con el nivel de contagio de esta enfermedad en ciertas zonas. Incluso serviría como referencia en tiempo real, es decir, un sistema que indique los niveles de contagio en ciertas zonas. Muchas otras aplicaciones pueden desarrollarse a partir de esta propuesta, por ejemplo, para la predicción de los niveles y aceleración del contagio de COVID-19 en zonas específicas de las ciudades.

A continuación, primero se describen algunas aplicaciones de la información contenida en las redes sociales, para luego describir algunos métodos para la inferencia de la ubicación geográfica de las publicaciones en redes sociales. Posteriormente y de manera más específica, se describe un modelo de aprendizaje computacional llamado *Transformadores* que identifica entidades de ubicación de las entradas de texto. Concluimos con un breve resumen del artículo y una discusión sobre algunas perspectivas de investigación.

Aplicaciones de las redes sociales

Una de las aplicaciones de la información contenida en las redes sociales que ha tenido mayor impacto, concierne a la predicción de resultados electorales. En [4], se describe cómo en las redes sociales la información puede impactar en la percepción de los votantes en una elección. Este hecho ha generado que los actores políticos busquen influenciar a sus posibles votantes de diversas maneras [7]. Por otro lado, los usuarios artificiales, también conocidos como *bots*, ya sean humanos o automatizados, representan una de las formas con mayor efecto en la manipulación de tendencias, incluidas las electorales. Recientemente, se han desarrollado diversos modelos para estudiar esta relación entre la información de las redes sociales y las tendencias electorales, por ejemplo, en [8] se propone un modelo estadístico de mediación basado en regresión. Estos fundamentan cuestiones como: el involucramiento de

los votantes en política, la participación con respecto a las elecciones, y su confianza en ellas.

Otras aplicaciones de reciente estudio, hacen referencia al flujo y tipo de información generada en los desastres naturales [5]. Se analizan diversos acontecimientos en los Estados Unidos de América: dos huracanes, Irene y Sandy; dos episodios de tornados y una inundación en el estado de Luisiana. Las publicaciones analizadas pertenecen a la red social *Twitter*. Algunas de las conclusiones de este estudio incluyen que los usuarios incrementan su actividad usual para compartir información relacionada antes y durante el acontecimiento. En [9], se describe como el análisis automático en esta red social puede ayudar en la detección de inundaciones.

Las redes sociales, en el ámbito de la salud, también han sido estudiadas. En [6], se menciona cómo la información que los usuarios publican puede ayudar en la identificación de ciertas patologías mentales, tales como, tendencias suicidas, depresión y ansiedad. Incluso, en [10] se han estudiado sistemáticamente diversos tipos de reacciones a ciertas fórmulas y medicamentos, a través de la información que los expertos intercambian en estas redes.

En la contingencia sanitaria producida por la enfermedad COVID-19, comenzaron a surgir estudios para revisar la calidad y veracidad de la información en los medios sociales. En [11] se retomaron algunas técnicas usadas en desastres naturales para identificar los contenidos de relevancia generados por la actual enfermedad. Es decir, se usó un método basado en Redes Neuronales (RN) para clasificar publicaciones en *Twitter* en tres categorías: recursos requeridos, disponibilidad de recursos, y otros. Esto también ayudó para planificar incluso medicinas.

En lo concerniente a este artículo se propone iniciar con la georreferenciación de las publicaciones en las redes sociales.

Localización geográfica de publicaciones

Algunos servicios de redes sociales proveen la ubicación geográfica de los usuarios y/o de las publicaciones. Existen investigaciones en donde los usuarios que permiten se conozca su ubicación son una minoría. Por ejemplo, en [12] se encontró un experimento con 19.6 millones de tuits, solamente el 0.7% tienen habilitada la opción de ubicación geográfica. Dada esta situación, se han desarrollado diversos métodos para inferir de ma-

nera automática la ubicación. Estos se pueden clasificar en tres tipos, de acuerdo a la información utilizada para realizar la inferencia. El primer tipo de métodos se basa en la información de los contactos, amigos o seguidores, según sea el caso, del usuario que hace la publicación [13]. El segundo tipo de métodos, se basa en la información de las publicaciones relacionadas, es decir, comentarios, respuestas, etcétera, según sea el tipo de red social [14]. El tercer tipo de métodos, se basa en la información contenida en la publicación misma [15]. Por lo general, se utilizan diversos mecanismos de inferencia basados en aprendizaje computacional, tales como, inferencia Bayesiana, redes neuronales, modelos de regresión, por mencionar algunos.

Los métodos de ubicación geográfica basados en la información contenida en las mismas publicaciones utilizan modelos para el procesamiento de textos sin estructura. Uno de los más empleados es el denominado Extracción de Entidades, también conocido como Reconocimiento de Entidades Nombradas (*Named Entity Recognition*, NER por sus siglas en inglés). Este modelo consiste en anotar entidades (objetos, personas, lugares, entre otros.) dentro de bloques de texto sin estructura, manejando categorías pre-definidas. Por ejemplo:

Daniel estudió en la Universidad Nacional hasta el año 2021.

Este texto se puede anotar de la siguiente manera:

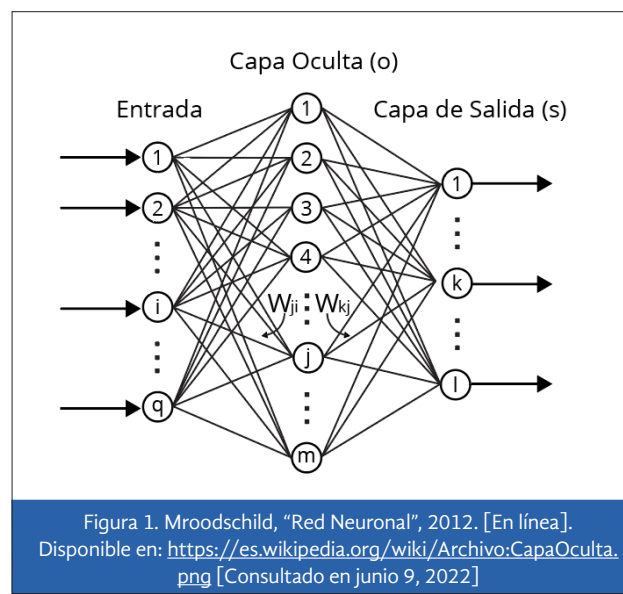
[Daniel]^{Persona} estudió en la [Universidad Nacional]^{Organización} hasta el [año 2021]^{Tiempo}.

En el texto anterior, se clasificó un componente léxico o *token*, Daniel, como persona; una organización de dos *tókenes* y una expresión temporal. Los sistemas NER modernos pueden ser tan precisos como personas expertas, por ejemplo, en [16] se describen modelos para el idioma inglés con una precisión de hasta el 92.8 por ciento.

Las arquitecturas utilizadas para el proceso de transducción (transformación), que implica la Extracción de Entidades (NER), de un texto sin anotar a un texto anotado¹, están basadas en Redes Neuronales (RN). Las RN son un método computacional que consiste en una serie de neuronas artificiales compuestas entre sí. A partir de ciertos datos de entrada, la RN computa datos de salida. En el caso de NER, los datos de entrada consisten de un texto sin anotar, y los datos de salida se conforma de un texto anotado. En la figura 1, se repre-

¹ Anotar entidades (objetos, personas, lugares, etc.) dentro de bloques de texto sin estructura utilizando categorías pre-definidas.

senta gráficamente una RN con una sola capa oculta, también conocida como capa intermedia. Cabe mencionar que las RN, comúnmente se integran por varias capas intermedias.

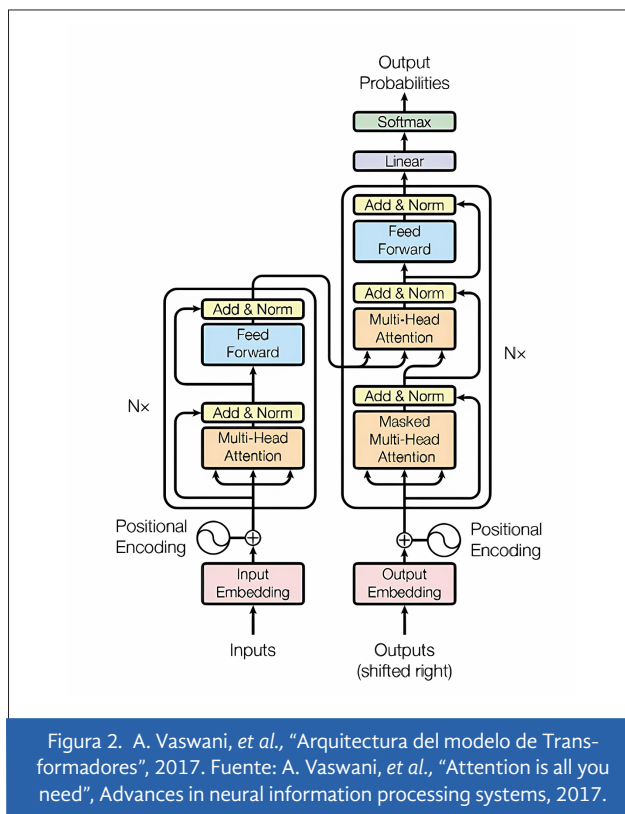


El entrenamiento de una RN es un proceso iterativo en el que la composición de las neuronas se ajusta para minimizar el error entre la salida que computa la RN con respecto a un conjunto de datos de entrenamiento. Entre más se entrene un RN, es decir, entre más datos tenga a su disposición, se vuelve más precisa.

La cantidad de capas intermedias de una RN y la cantidad de datos de entrenamiento son los principales factores en el costo de cómputo, tanto de memoria como de tiempo. Una de las técnicas de optimización más útiles en los procesos de cómputo, consiste en su paralelización.

Geolocalización de COVID-19 con Transformadores

Los *Transformadores* son un modelo de aprendizaje computacional basado en RN. La característica principal se centra en el concepto de atención, el cual permite hacer ajustes particularmente enfocados a la composición de las neuronas durante el proceso de entrenamiento de la RN. Un caso exitoso basado en *Transformadores* es el llamado *Representación Bidireccional de Codificadores de Transformadores*, BERT por sus siglas en inglés [17]. Este modelo propone la paralelización de ciertas tareas durante el proceso de entrenamiento de la RN. En la figura 2, se representa gráficamente la arquitectura de este modelo.



En el idioma inglés, el modelo BERT ha logrado una precisión de hasta el 92.8% para la tarea de Extracción de Entidades (NER) [16]. Para el idioma castellano, hay estudios que concluyen hasta un 88.43% de precisión para NER [18].

En el contexto de publicaciones en redes sociales, en [19] se aplica el modelo BERT para NER en castellano [18]. Esta prueba, concretamente, se llevó a cabo en *Twitter*, en donde la NER reportó una precisión de 93%. Sin embargo, la extracción de las entidades relacionadas con la ubicación no fue suficiente para determinar si la publicación se realizó en la ubicación extraída. Por lo que en [19] también se reportó la recolección de 220,426 publicaciones en *Twitter* que contaban con la ubicación habilitada. Las entidades extraídas concernientes a la ubicación coincidieron en un 29 por ciento.

Adicionalmente, al experimento NER descrito en [19] con el modelo BERT, se realizó otro sondeo para detectar publicaciones en *Twitter* sobre COVID-19. En concreto, se buscaron publicaciones con las palabras "COVID" y "contagio". Se analizaron 104,035 tuits, de los cuales 15,050 estaban escritos en idioma castellano.

De estas últimas publicaciones, se detectaron 3,369 publicaciones relacionadas con alguna ubicación geográfica. Finalmente, el algoritmo detectó 15 publicaciones sobre COVID-19 con la ubicación geográfica correctamente. Este resultado se obtuvo de un total 76 tuits con la ubicación geográfica habilitada.

Conclusiones

De manera general, se abordaron casos y modelos específicos en las redes sociales, tales como: predicciones electorales, detección de patrones durante desastres naturales, y la identificación de enfermedades. Este artículo presentó un método para analizar información que tuviera algún vínculo con la contingencia sanitaria producida por la enfermedad COVID-19 y así relacionar las publicaciones con determinadas ubicaciones geográficas. De esta manera se identificó el nivel de contagio en zonas específicas. La ubicación de estas notas fue una de las características de mayor relevancia.

Como revelan algunos estudios, aunque varios de los servicios de redes sociales permiten compartir la ubicación tanto de los usuarios como de las publicaciones, solo una minoría de los usuarios lo ponen en práctica. Por lo que el desarrollo de métodos para inferir su ubicación toma evidente relevancia. De manera particular, el método basado en la Extracción de Entidades (NER) de ubicación, describe uno de los algoritmos más exitosos para dicha tarea, como lo es el basado en Transformadores (BERT). Los experimentos antes mencionados concluyen que, aunque la extracción de la ubicación en el texto se puede hacer de manera muy precisa, cuando se detectan palabras relacionadas con alguna ubicación geográfica, el nivel de precisión disminuye.

Como perspectiva de investigación inmediata, se tiene el estudio de métodos híbridos para la inferencia de la ubicación geográfica en redes sociales, como el reportado en [20]. Este tipo de métodos, además de considerar el texto, también considera la información de reenvíos, respuestas y referencias. La información de los usuarios también puede ser analizada para el mismo propósito de inferencia de la ubicación.

Agradecimientos

Investigación realizada gracias al Programa UNAM-PA-PIIT IA104122, IA105420, IA105320 y TA10102.

BiBliogr Afí A

- [1] "Number of social network users worldwide 2017 to 2025," *Statista*, 2021. [En línea]. Disponible en: <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>. [Consultado en abril 27, 2022].
- [2] "2020 World Population Data Sheet," *World Population Data Sheet*, 2021. [En línea]. Disponible en: <https://interactives.prb.org/2020-wpds/> [Consultado en abril 27, 2022].
- [3] "Censo de población y vivienda 2020," *Instituto Nacional de Geografía, Estadística e Informática (INEGI)*, 2020. [En línea]. Disponible en: <https://www.inegi.org.mx/programas/ccpv/2020/> [Consultado en abril 27, 2022].
- [4] P. Metaxas y E. Mustafaraj, "Social Media and the Elections," *Science*, vol. 338, no. 6106, 2012.
- [5] M. T. Niles, B. F. Emery, A. J. Reagan, P. S. Dodds y C. M. Danforth, "Social media usage patterns during natural hazards," *PLOS ONE*, vol. 14, no. 2, 2019. DOI:[10.1371/journal.pone.0210484](https://doi.org/10.1371/journal.pone.0210484)
- [6] S. Chancellor y M. De Choudhury, "Methods in predictive techniques for mental health status on social media: a critical review," *npj Digital Medicine*, vol. 3, no. 43, 2020. <https://doi.org/10.1038/s41746-020-0233-7>
- [7] A. J. Stewart, M. Mosleh, M. Diakonova, A. A. Arechar, D. G. Rand y J. B. Plotkin, "Information gerrymandering and undemocratic decisions," *Nature*, vol. 573, 2019.
- [8] H. Lee, "Voters' involvement, attitude, and confidence in the era of new media," *Palgrave Communications*, vol. 6, no. 1, 2020.
- [9] K. Shoyama, Q. Cui, M. Hanashima, H. Sano y Y. Usuda, "Emergency flood detection using multiple information sources: Integrated analysis of natural hazard monitoring and social media data," *Science of The Total Environment*, vol. 767, 2021.
- [10] C. C. Yang, H. Yang, L. Jiang y M. Zhang, "Social Media Mining for Drug Safety Signal Detection," de *International Workshop on Smart Health and Wellbeing*, Maui, Hawaii, 2012.
- [11] S. Behl, A. Rao, S. Aggarwal, S. Chadha y H. Pannu, "Twitter for disaster relief through sentiment analysis for COVID-19 and natural hazard crises," *International Journal of Disaster Risk Reduction*, vol. 55, 2021.
- [12] M. Graham, S. A. Hale y D. Gaffney, "Where in the World Are You? Geolocation and Language Identification in Twitter," *The Professional Geographer*, vol. 66, no. 4, 2014.
- [13] A. Sadilek, H. Kautz y J. P. Bigham, "Finding your friends and following them to where you are," de *ACM international conference on Web search and data mining*, Seattle, Washington, 2012.
- [14] W. Li, P. Serdyukov, A. P. d. Vries, C. Eickhoff y M. Larson, "The where in the tweet," *ACM international conference on Information and knowledge management*, Glasgow, Scotland, 2011.
- [15] H.-W. Chang, D. Lee, M. Eltaher y J. Lee, "@Phillies Tweeting from Philly? Predicting Twitter User Locations with Spatial Word Usage," de *International Conference on Advances in Social Network Analysis and Mining*, Istanbul, 2012.

[16] T. Young, D. Hazarika, S. Poria y E. Cambria, "Recent Trends in Deep Learning Based Natural Language Processing," *IEEE Computational Intelligence Magazine*, vol. 13, no. 3, 2018.

[17] A. Vaswani, *et al.*, "Attention is all you need," de *Advances in neural information processing systems*, 2017.

[18] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang y J. Pérez, "Spanish Pre-Trained BERT Model," de *Practical ML for Developing Countries Workshop at ICLR*, 2020.

[19] A.-D. Ambrosio-Aguilar, E. Bárcenas, G. Molero-Castillo y R. Aldeco-Pérez, "Geolocation of Tweets in Spanish with Transformer Encoders," de *International Conference in Software Engineering Research and Innovation*, 2021.

[20] B. Huang y K. M. Carley, "A Hierarchical Location Prediction Neural Network for Twitter User Geolocation," de *Conference on Empirical Methods in Natural Language Processing*, 2019.

Fecha de recepción: noviembre 23, 2021

Fecha de publicación: marzo de 2023

Cómo se cita

E. Bárcenas Patiño, A.D. Ambrosio Aguilar, G.G. Molero Castillo, R. Aldeco Pérez, "Análisis de redes sociales en torno a la emergencia sanitaria de COVID-19". *TIES, Revista de Tecnología e Innovación en Educación Superior*, no. 7, marzo, 2023. [En línea]. Disponible en: <https://ties.unam.mx/> [Consultado en mes día, año].