ISSN 1870-4069

Detección de noticias falsas en sitios web y redes sociales: Una investigación del estado del arte

Sergio Damián, Alexander Gelbukh, Hiram Calvo

Instituto Politécnico Nacional,
Centro de Investigación en Computación,
México

b190394@sagitario.cic.ipn.mx, gelbukh@gelbukh.com,
hcalvo@cic.ipn.mx.

Resumen. La divulgación de noticias falsas en redes sociales ha provocado grandes impactos en la sociedad durante los últimos años, por lo que es importante evitar su propagación. En la actualidad, existen diferentes formas de atacar este problema, las cuales parten por la detección de estas noticias e incluso perfilar a los usuarios que tienden a divulgar este tipo de noticias, ya sea de forma intencional o no. El presente artículo discute las características que posee una noticia para ser catalogada como "falsa", algunas propuestas de solución en el estado del arte, para finalmente concluir sobre los resultados obtenidos y sugerir posible trabajo a futuro en el área. Todo esto con un enfoque en las redes sociales y sitios de internet dedicados a la divulgación de noticias.

Palabras clave: Noticias falsas, aprendizaje automático, desinformación, detección, clasificación

Fake News Detection in Web Sites and Social Media: A Survey of the State of the Art

Abstract. Fake news spreading in social media has caused a big impact in society during the last years. Nowadays, there are different ways to attack this problem, some of them start by detecting fake news and other ones by profiling the users who tend to spread this kind of news, either intentionally or not. The present paper discusses the features and characteristics that can represent what fake news are, some solution proposals in the state of the art, and finally give some conclusions about the obtained results and suggest possible future work in the area. All of this focusing in social media and web sites that are dedicated to spread news.

Keywords: Fake news, automatic learning, misinformation, detection, classification.

1. Introducción

De acuerdo con la investigación en [1] el impacto de la divulgación de noticias por parte de reporteros y periodistas ha sido un factor muy importante en la sociedad, ya

que tiende a ser la principal fuente de consumo de información fiable por parte de los ciudadanos. No obstante, hoy en día se manejan las redes sociales como un medio de difusión más eficiente que los medios convencionales como la radio y el periódico, por lo que se permite que cualquier usuario sea capaz de compartir información ya sea verídica o engañosa. Es por este motivo que la detección y la prevención del consumo de noticias en redes sociales, presenta un problema a resolver con el fin de proteger el consumo de información fiable y verídica.

Podemos notar con el paso del tiempo, diferentes acontecimientos en donde se comenta que la gente tiende a razonar incorrectamente o a cambiar de opinión debido a información falsa o engañosa. Un ejemplo claro de esto es el rumor descrito en [2] en donde un ciudadano arremetió en un restaurante con un arma de fuego, argumentando que leyó información de que en ese lugar existía abuso infantil, liderado por la excandidata a la presidencia en Estados Unidos en 2016, Hillary Clinton. Por supuesto que esto no se limita únicamente al ámbito político, pero principalmente el auge y la comprensión de la gravedad del problema comenzó con este tipo de noticias durante esas elecciones presidenciales.

Definir el concepto de noticias falsas, (del inglés *fake news*), tiene un amplio panorama debido a las diferentes características que éstas pudieran poseer, desde el motivo o la intención por la que se está difundiendo ese tipo de información, hasta su estilo de escritura, e incluso la desinformación que el usuario tiene al momento de escribir y/o compartir algún hecho en las redes sociales. La definición que nos ofrece el sitio Web Wikipedia en español [3] es: "Las noticias falsas son un tipo de bulo que consiste en un contenido pseudo periodístico difundido a través de portales de noticias, prensa escrita, radio, televisión y redes sociales y cuyo objetivo es la desinformación". En el trabajo realizado en [4], se muestra una investigación donde concluyen que normalmente el generalizar el concepto tiende a ser de mucha dificultad, debido a la amplia gama de características que pudiesen definirlo. No obstante, nos ofrece la siguiente definición: "Las noticias falsas se refieren a todo tipo de historia falsa o noticia que principalmente está publicada y distribuida en Internet, con el fin de engañar a propósito o atraer lectores con objetivos financieros, políticos, entre otros".

Finalmente, podemos encontrar que las noticias falsas no son más que una categoría de publicación de información falsa, dentro de las cuales se encuentran: Rumores, teorías de conspiración, sátira, desinformación, propaganda, entre otras. [5].

Con ayuda de estas definiciones, en este trabajo trataremos la definición como todas aquellas publicaciones de contenido engañoso que pueda afectar la opinión o el punto de vista respecto a algún tema en el lector.

El presente trabajo presenta un análisis de diez proyectos desarrollados en los últimos dos años [9-11, 14-20], los cuales están enfocados en la detección de noticias falsas (de acuerdo con su propia definición) en redes sociales y sitios web dedicados a la recolección de noticias utilizando diferentes metodologías, algoritmos y/o técnicas que se conocen en la actualidad y que tienden a presentar buenos resultados. Es así como el presente trabajo se organiza de la siguiente manera: en la sección 2 se discute sobre cómo se recolecta la información de las publicaciones y cómo se preprocesa; en la sección tres se describen las diferentes características del texto que el estado del arte considera para su extracción y evaluación; en la sección cuatro se describen las técnicas y algoritmos utilizados para la solución del problema; y finalmente, en la sección cinco

se discute sobre cómo se ha abordado el problema, la dirección a donde se considera que la investigación se dirigirá en el futuro y conclusiones adicionales.

2. Recolección del conjunto de datos y preprocesamiento

Categorizar una publicación como noticia falsa, es un reto que conlleva a cuestionarnos en primer lugar sobre qué se considera como noticia falsa y qué es una noticia verídica. En [6] se menciona que diversos autores consideran a las noticias falsas como una subcategoría de lo que se considera como "información engañosa", la cual también puede ser catalogada como rumores, spam, etc., por lo que es importante que en cada trabajo se determine qué es lo que se considerará para la creación de su conjunto de datos y su detección. No obstante, existen algunos ejemplos de conjuntos de datos previamente catalogados, tal es el caso de la red social de Twitter.

Dentro de las propuestas de solución desarrolladas para la detección de noticias falsas, se ha generalizado dos grupos de soluciones, unas basadas en el contenido y otras basadas en la red que se está investigando [6-8]. No obstante, independientemente de qué tipo de propuesta de solución es realizada, se determina que se requiere un preprocesamiento del conjunto de datos, una representación y una selección de algoritmos, ya sean de aprendizaje automático clásicos o de aprendizaje profundo para el desarrollo de una solución. Por supuesto existen algunas otras propuestas un tanto alejadas de un desarrollo con algoritmos de aprendizaje automático, como pueden ser detección de anomalías, patrones difusos, etcétera [6]. Sin embargo, el enfoque de este trabajo consiste en analizar las propuestas que se han desarrollado para la detección y clasificación de noticias especialmente mediante el uso de algoritmos de aprendizaje automático y de aprendizaje profundo.

La tarea de recolectar información y catalogar si una noticia es falsa o no, conlleva una complejidad alta, debido a que se debe de tener conocimiento de la información presentada en el conjunto de datos, por lo cual, la gran mayoría de las soluciones analizadas tienden a trabajar con datos previamente categorizados de fuentes de bases de datos en internet. Sin embargo, el preprocesamiento de éstos es la parte en la que cada solución empieza a diferir una de otra. La tabla 1 muestra ejemplos de conjuntos de datos utilizados por algunas soluciones de detección de noticias falsas [9, 10, 15, 17]. En adición, se ha encontrado en el estado del arte que también es posible crear un conjunto de datos propio mediante APIs que son procedentes de los sitios web o redes sociales, por ejemplo, el uso de Twitter API para la extracción de publicaciones y sus correspondientes metadatos [16].

Tabla 1. Ejemplos de conjuntos de datos utilizados.

Nombre del conjunto de datos	**Proveedor**
BuzzFeed Political News dataset	BuzzFeed
LIAR dataset	Polifact
real_or_fake news dataset	Kaggle
Random Political News dataset	Data.world

Durante el preprocesamiento, es común realizar una limpieza a los datos, donde se utilizan técnicas de NLP (Por sus siglas en inglés, procesamiento de lenguaje natural)

como convertir los textos a letras minúsculas, eliminar puntuación o remover stop words o palabras vacías, con el fin de estandarizar la información para un análisis más eficiente. En [11,15] se utiliza este tipo de preprocesamiento, donde además se extraen características adicionales que se encuentran de manera implícita en los textos, las cuales se discutirán en la sección 4. Además, la representación final del conjunto de datos puede ser implementada utilizando alguna técnica como *Bag of words* [10], *Count Vectorizer* [10,16] o TF-IDF [10,14–16].

3. Extracción de características en las noticias falsas

Durante el desarrollo de una propuesta de solución que deriva en utilizar algoritmos de aprendizaje automático o aprendizaje profundo, se extraen algunas características de los textos, (las cuales son mencionadas en las secciones 3.1 y 3.2), como una fuente de información adicional para mejorar el desempeño del algoritmo. A continuación, se describen en qué consisten las características mencionadas en las publicaciones [6–8] las cuales fueron clasificadas en dos tipos: basadas en el contenido y basadas en el contexto.

3.1. Características basadas en el contenido

Este tipo de características consisten principalmente en analizar a fondo la redacción que tiene cada publicación, para determinar si existen características adicionales implícitas que pudieran permitir una mejor detección de noticias falsas. Por lo tanto, éstas son extraídas cuando se realiza un análisis léxico, sintáctico o semántico. En la tabla 2 se muestran algunos ejemplos de características basadas en el contenido y de qué tipo son, utilizadas en [11, 15, 19].

Tabla 2. Lista con ejemplos de características basadas en el contenido.

Característica	Tipo de característica
Uso de N-gramas para palabras y caracteres	Sintáctico
Conteo de signos de puntuación	Léxico
Uso de letras mayúsculas	Léxico
Uso de palabras altisonantes	Semántico
Hashtags o emoticones (sentimientos)	Semántico

3.2. Características basadas en el contexto

Este tipo de características consisten principalmente en analizar información referente al tipo de usuarios que están difundiendo, leyendo y/o compartiendo las noticias en una cierta plataforma social en internet. También incluye metadatos del sitio web como el analizar la propagación de una cierta publicación en el medio. En [13] es mencionada la importancia de obtener información de los usuarios, como su edad, género, procedencia, etc.

Tabla 3. Ejemplos de características basadas en el contexto.

Característica	Tipo de característica
Origen geográfico del usuario	Usuario
Propagación de una publicación	Red
Número de páginas visitadas	Usuario
Número de “me gusta”, “retweets”, etc.	Usuario
Número de publicaciones compartidas	Usuario
Red de amigos	Red
Comentarios en una publicación	Red

Tabla 4. Algoritmos de aprendizaje máquina.

Algoritmo	Porcentaje de trabajos analizados donde se utiliza
Decision Tree	20%
Logistic Regression	30%
SVM	40%
Naive Bayes	20%
Random Forest	40%
Gradient Boosting	30%

En la tabla 3 se muestran algunos ejemplos de características basadas en el contexto, definidas en [13, 20].

En la tabla 4 se presentan los algoritmos utilizados en los trabajos que se orientaron en el aprendizaje automático [9-11,14-18].

4. Métodos y algoritmos utilizados para la solución del problema

Como se menciona en la sección dos, en esta sección se analizan las soluciones propuestas en el estado del arte basadas en el aprendizaje máquina y en el aprendizaje profundo exclusivamente. Cabe mencionar que en el estado del arte, existen trabajos que no se limitan a trabajar con un solo tipo de algoritmo, si no que parten de desarrollos de aprendizaje automático clásicos, para posteriormente realizar una comparación de resultados con algoritmos de aprendizaje profundo u otro tipo de modelos. Así también, el desarrollo no se limita a trabajar con el texto de las publicaciones, sino también algunos trabajos consideran trabajar con las imágenes que acompañan a los textos.

4.1. Soluciones basadas en aprendizaje máquina

Dado que podemos catalogar este problema como una clasificación, existen diversos algoritmos que permiten identificar las características más importantes y orientar el

Tabla 5. Técnicas de aprendizaje profundo.

Algoritmo	Porcentaje de trabajos analizados donde se utiliza
MLP (Multilayer Perceptron)	10%
CNN (Convolutional Neural Net)	20%
LSTM (Long Short-Term Memory)	10%
Bi-LSTM (Bidirectional LSTM)	10%
GCN (Graph Convolutional Net)	20%

modelado de la solución a partir de éstas. Por lo tanto, es común observar que existen diversas soluciones basadas en algoritmos de aprendizaje automático.

Como se puede observar, los algoritmos de SVM y Random Forest son los más populares para utilizarlos en la detección de noticias falsas de acuerdo con los trabajos analizados en el estado del arte (en ambos casos, el 40% de los trabajos empleó estos algoritmos), los cuales, a su vez, también han obtenido buenos resultados en comparación con los obtenidos en trabajos enfocados en aprendizaje profundo.

4.2. Soluciones basadas en aprendizaje profundo

Estas soluciones tienden a estar acompañadas de técnicas como *Word Embeddings*, ya sea que se utilice Doc2Vec [16], GloVe [14,18,19] o algún *embedding* personalizado. La tabla 5 presenta las técnicas de aprendizaje profundo utilizadas en el estado del arte.

Se puede observar que las redes neuronales convolucionales son las preferidas para la implementación de la solución, de acuerdo con el estado del arte analizado.

5. Conclusiones y trabajo a futuro

La detección de noticias falsas en sitios web y redes sociales ha sido un problema importante a resolver en la actualidad, por lo cual, el presente trabajo ofrece un panorama actual de las diferentes tecnologías de aprendizaje automático y aprendizaje profundo para la implementación de posibles soluciones. Una vez que se han analizado las diferentes metodologías y propuestas de solución en el estado del arte, es importante mencionar que la selección de una característica, técnica o modelo, depende mucho del tipo de información que se analizará, así como también de factores como los recursos computacionales con los que se cuenta.

Se analizó que existen diferentes fuentes de datos de donde podemos partir para la elaboración de una propuesta de solución, algunos de extracción de características y los diferentes algoritmos de aprendizaje automático y aprendizaje profundo, esto con la finalidad de presentar algunas opciones que han resultado adecuadas para el problema y que se pueda partir desde este tipo de trabajos hacia propuestas de solución más elaboradas y con mejor adaptación al detectar si una publicación es engañosa o no.

Puede ser interesante considerar otro tipo de técnicas y también si existen autores que hayan desarrollado su propio clasificador para atacar este problema con mayor eficacia, de acuerdo a las características propias que se presentan.

Referencias

1. Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M., Procter, R.: Detection and resolution of rumours in social media: a survey. ACM Comput. Surv., 51(2), pp. 32:1–32:36 (2018)
2. Kang, C., Goldman, A.: Washington pizzeria attack. Fake News Brought Real Guns, 5 (2016)
3. Wikipedia: Fake news. https://es.wikipedia.org/wiki/Fake_news (2020)
4. Zhang, X., Ghorbani, A.A.: An overview of online fake news: Characterization, detection, and discussion. Information Processing and Management (2020)
5. Meel, P., Vishwakarma, D.K.: Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. Expert Systems with Applications (2020)
6. Bondielli, A., Marcelloni, F.A.: Survey on fake news and rumour detection techniques. Information Sciences (2019)
7. Conroy, N.J., Rubin, V.L., Chen, Y.: Automatic deception detection: methods for finding fake news. In: Proceedings of the Association for Information Science and Technology (2015)
8. Zhou, X., Zafarani, R.; Network-based fake news detection. ACM SIGKDD Explorations Newsletter (2019)
9. Ozbay, F.A., Alatas, B.: Fake news detection within online social media using supervised artificial intelligence algorithms. Physica A: Statistical Mechanics and Its Applications (2020)
10. Vasuagarwal, H., Sultana, P., Malhotra, S., Sarkar, A.: Analysis of classifiers for fake news detection. In: International Conference on Recent Trends in Advanced Computing (2019)
11. Aldwairi, M., Alwahedi, A.: Detecting fake news in social media networks. Procedia Computer Science (2018)
12. Reis, J.C.S., Correia, A., Murai, F., Veloso, A., Benevenuto, F.: Explainable machine learning for fake news detection. In: Proceedings of the 11th ACM Conference on Web Science (2019)
13. Shu, K., Zhou, X., Wang, S., Zafarani, R., Liu, H.: The role of user profiles for fake news detection. In: Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (2019)
14. Katsaros, D., Stavropoulos, G., Papakostas, D.: Which machine learning paradigm for fake news detection? In: Proceedings - IEEE/WIC/ACM International Conference on Web Intelligence, WI (2019)
15. Hnin-Ei, W., ZarZar, W.: Content based fake news detection using n-gram models. In: Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services (iiWAS) ACM, Munich, Germany, 5 (2019)
16. Helmstetter, S., Paulheim, H.: Weakly supervised learning for fake news detection on Twitter. In: Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (2018)
17. Rasool, T., Butt, W.H., Shaukat, A., Akram, M.U.: Multi-label fake news detection using multi-layered supervised learning. In: ACM International Conference Proceeding Series (2019)
18. Benamira, A., Devillers, B., Lesot, E., Ray, A.K., Saadi, M., Malliaros, F.D.: Semi-supervised learning and graph neural networks for fake news detection. In: Proceedings of

the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) Association for Computing Machinery, pp. 568–569 (2019)
19. Abedalla, A., Al-Sadi, A., & Abdullah, M. A closer look at fake news detection: A deep learning perspective. In: ACM International Conference Proceeding Series. https://doi.org/10.1145/3369114.3369149 (2019)
20. Reis, J. C. S., Correia, A., Murai, F., Veloso, A., & Benevenuto, F. Explainable machine learning for fake news detection. In: WebSci 2019, Proceedings of the 11th ACM Conference on Web Science. https://doi.org/10.1145/3292522.3326027 (2019)