REVISTA INTERNACIONAL DE INVESTIGACIÓN CIENTÍFICA Y TECNOLÓGICA VOLUMEN 9, NÚMERO 03, MARZO DE 2020

Una encuesta sobre la detección de noticias falsas en las redes sociales Medios que utilizan redes neuronales profundas

L. Alekya, L. Lakshmi, G. Susmitha, S. Hemanth

Abstracto: Debido a las tecnologías emergentes y al crecimiento de la población, la tasa de uso de las redes sociales ha aumentado rápidamente. A partir de ahora, las redes sociales se han convertido en las noticias diarias del mundo en lugar de los canales de noticias y los periódicos. La difusión de noticias falsas puede causar estragos. Y lo más probable es que estemos difundiendo las noticias falsas a nuestro entorno mediante la confianza digna en las redes sociales que finalmente se está engañando con demasiada confianza. En los próximos días, no podremos distinguir entre noticias reales y noticias falsas. La tasa de noticias falsas se ha convertido en un disfraz. En este proceso, hemos realizado encuestas para detectar noticias falsas utilizando técnicas y algoritmos de Deep Learning. Hemos probado algunos de los conjuntos de datos como LIAR, Buzzfeed, PolitiFact, Kaggle y McIntire. En este artículo para detectar las noticias falsas, hemos comparado algoritmos de aprendizaje automático como Naïve-Bayes ', SVM, Decision Tree, AdaBoost, etc. La comparación de la precisión obtenida con vainilla y LSTM son menos precisas que GRU y CNN. Buscamos aumentar la precisión aplicando un modelo híbrido entre las técnicas GRU y CNN en el mismo conjunto de datos.

Términos del Índice: Redes neuronales convolucionales, Unidad recurrente cerrada, Máquina de vectores de soporte, LSTM, Máquina de vectores de soporte, Redes neuronales recurrentes, Bolsa de palabras.

1. INTRODUCCIÓN

ন্দ্ৰে Traducido del inglés al español - www.onlinedoctranslator.com

Las noticias falsas son una de las principales intimidaciones en el periodismo y La dependencia a corto plazo se produce debido a la desaparición democracia. Las noticias falsas hicieron que la gente pudiera tener un problema de gradiente. Este problema de degradado que desaparece afecta a la no diferenciación de las noticias reales y las noticias falsas. Superar las redes feed forward y las redes recurrentes. Posteriormente, la alimentación se realiza mediante el uso de algunos sitios web, herramientas y plataformas en las que se entrenan las redes de reenvío, se desarrollan y se convierten en muy introducidas para detectar la diferenciación de las noticias reales y las redes de reenvío profundas falsas. Luego de desdoblarlos una nueva capa de noticias [1]. Los farsantes crean una noticia que hace que se genere. Hay una solución para resolver la desaparición de la mente de los lectores de Gradient desesperados por saber qué hay en las noticias por problema que es mediante la versión actualizada del modelo RNN. Gated viendo el título de una noticia en particular. Que las mentes desesperadas son Recurrent Unit es la versión actualizada del modelo RNN. Esto les da la oportunidad a los falsificadores de crear una unidad Recurrente cerrada tan malsana que usa puertas vectoriales para resolver este problema. Noticias. Aunque la detección de noticias falsas no está completamente usando UPDATE GATE y RESETGAT, estos dos vectores gobiernan el mundo pero se está generando mucho interés [2]. Además de decidir qué salida se debe transmitir y mantener con las redes sociales, hay algunos sitios web que difunden información durante un período más larqo. La unidad Gated Recurrent no falsifica noticias, como The Gateway Pundit, Gummy post, etc., requiere unidades de memoria, mientras que las de vainilla y LSTM lo son. Hemos probado en LSTM donde el LSTM es más caro y se necesita estos dos vectores gobiernan el mundo pero se está generando mucho interés [2]. Además de decidir qué salida se debe transmitir y mantener con las redes sociales, hay algunos sitios web que difunden información durante un período más largo. La unidad Gated Recurrent no falsifica noticias, como The Gateway Pundit, Gummy post, etc., requiere unidades de memoria, mientras que las de vainilla y LSTM lo son. Hemos probado en LSTM donde el LSTM es más caro y se necesita estos dos vectores gobiernan el mundo pero se está generando mucho interés [2]. Además de decidir qué salida se debe transmitir y mantener con las redes sociales, hay algunos sitios web que difunden información durante un período más largo. La unidad Gated Recurrent no falsifica noticias, como The Gateway Pundit, Gummy post, etc., requiere unidades de memoria, mientras que las de vainilla y LSTM lo son. Hemos probado en LSTM donde el LSTM es más caro y se necesita. Para modificar los encabezados en ejecución de GNN seleccione Ver, para calcular la salida de la red, la optimización y también para aplicar Encabezado y Pie de página. Haga clic dentro del cuadro de texto para escribir el nombre de la retropropagación. La celda de memoria envía la entrada a los nodos de la revista a la que se envía el artículo y el manuscrito

(convencional oculto) [3]. Cuando agregamos memoria externa ese número de identificación. Hacer clic en la flecha hacia adelante en la herramienta emergente aumenta el peso de los nodos asociados con los sitios web. barra para modificar el encabezado o pie de página en las páginas siguientes. En ese caso, nuevamente necesitamos entrenar los nodos. Esto es muy

difícil y la dimensionalidad de los nodos aumenta y 2. REVISIÓN DE LA LITERATURA

no se puede lograr la optimalidad. Las redes neuronales convolucionales Estee Van Der Walt [5] centradas en la detección de noticias falsas se utilizan principalmente en la clasificación de textos. La extensión de Natural creada por bots y humanos. Los bots no dieron el mejor procesamiento del lenguaje en las redes neuronales convolucionales. Comparamos los resultados con las cuentas humanas. Las características solían decir que es la mejor herramienta para extraer características. Si bien la detección de cuentas falsas son modelos de aprendizaje automático supervisados que estudian las dependencias a largo plazo, no es eficaz. El anhelo de entrenar las características de ingeniería. El trabajo futuro de este proyecto las dependencias de término se superaron mediante la técnica RNN. Se trata de mejorar los conjuntos de funciones. Las cuentas creadas por bots RNN vuelven a modelar hay un problema con Vanilla que los leads no pueden competir con las cuentas creadas por humanos; Vicio

a dependencias a corto plazo. Esta dependencia a corto plazo, viceversa, las cuentas creadas por bots no son similares a las de los humanos, ocurre cuando la información obtenida no puede durar más cuentas. Se ha generado F1score para las cuentas falsas tiempo. Este modelo se utiliza para elegir funciones de contenido creados por humanos son aproximadamente el 49%. Las nuevas func

automáticamente [4].

La Sra. L. Alekya actualmente cursa una maestría en el Instituto de Tecnología MLR, Hyderabad India, PH-7093688636.

Correo electrónico: <u>alekya.lingasetty09@gmail.com</u>

La Dra. L.Lakshmi trabaja actualmente como profesora en BVRIT Hyderabad College of Engineering for Women, Hyderabad India, PH-9010482007. Correo electrónico: <u>laxmi.slv@qmail.com</u>

El Dr. G.Susmitha trabaja actualmente como profesor en el Instituto de Tecnología MLR, Hyderabad India, PH-9885211511.

Correo electrónico: <u>s usmita@mlrinstitutions.ac.in</u>

El Sr. S. Hemanth actualmente está cursando una licenciatura en el Instituto de Tecnología MLR, Hyderabad India, PH-7093688636.

creados por humanos son aproximadamente el 49%. Las nuevas funciones que se encuentran en las SMP mostrarán mejores resultados. Marco L. Della Vedova [6] propuso el método automático de detección de noticias falsas para combinar métodos sociales y métodos basados en contenido. El desempeño de los métodos sociales no fue bueno. Cuando se combinan los métodos sociales y los métodos basados en contenido, han dado el mejor índice de precisión en la búsqueda de artículos de noticias. Se introdujo el método HC-CB-3 para ofrecer una alta precisión. El uso del método basado en contenido ha reducido la difusión de noticias falsas. Además, introduzca los bots de chat en el entorno de mensajería de Facebook para detectar noticias reales y falsas. Esto se logró con un 81%. El trabajo futuro es capacitar a los bots de chat con respecto a los conjuntos de datos a qué país pertenecen las noticias y eliminar las noticias italianas.

Liang Wu [7] afirmó que la clasificación de mensajes se utilizó para encontrar los mensajes que se difunden en las redes sociales que no son apropiados, confiables, etc. Y propuso a Trace Miner clasificar los mensajes en las redes sociales, con esto un LSTM-RNN de extremo a extremo. También se incluyó el modelo de clasificación. Trace Miner ha dado buenos resultados en conjuntos de datos insuficientes. El trabajo futuro se realizará en tareas de minería de red utilizando

De acuerdo con el enfoque de Hadder Ahmed [8] sobre las noticias falsas que se detectaron utilizando el análisis de Ngram, se describió cómo las noticias falsas se han transmitido a las redes sociales y que las noticias falsas se han difundido ampliamente entre los jóvenes, atrajo la atención de los jóvenes para creer solo las noticias de las redes sociales, no las noticias impresas. Usando alguna extracción de características

Tabla 1: Tabla comparativa de técnicas de PNL

Técnicas de PNL	Descripción		
Bolsa de palabras	Este modelo cuenta el número de		
	veces que la palabra apareció en el		
	texto.		
Término frecuencia inversa	Tf-Idf dice la importancia de la		
Frecuencia del documento	palabra en una oración. Poniendo		
	peso a cada palabra		
Guante	Este mapa los palabras		
	significativamente		
Word2Vec	Esta es una incrustación de dos capas		
	que están capacitadas para reconstruir		
	palabras de contenido lingüístico.		

técnicas, e investigado en diferentes extracción de características 4 ALGORITMOS DE APRENDIZAJE MÁQUINA UTILIZADOS EN y técnicas de aprendizaje automático. El modelo propuesto tal DETECCIÓN DE FALSAS NOTICIAS

como característica unigram y clasificador SVM ha alcanzado el 92% precisión. XinYi Zhou [9] ha propuesto detectar noticias falsas en el contenido de las redes sociales. Los modelos han logrado una precisión

4.1 Clasificador ingenuo de Bayes (modelo de aprendizaje generativo)

Naïve Bayes es un modelo probabilístico de aprendizaje automático y con aproximadamente un 80-88% en la aplicación basada en contenido, este es un clasificador probabilístico que hace clasificaciones utilizando detección de noticias falsas híbrida y basada en propagación. Con regla de decisión menos max posteriori. El clasificador ingenuo de Bayes se deriva de la cantidad de conocimientos que este modelo ha proporcionado con buena precisión, a partir del teorema de Bayes que se utiliza en las tareas de clasificación. Estos tres de ellos el modelo propuesto funcionó bien para predecir el clasificador es escalable y el algoritmo tradicional de elección y las noticias falsas. Las noticias falsas han creado una sensación de lo que se relacionan con las aplicaciones del mundo real. Estos del mundo real noticias reales. La noticia se ha representado con breves palabras, con aplicaciones que dan una rápida respuesta a los usuarios. Este algoritmo menos contenido. El trabajo futuro es utilizar nuevas relaciones y se utiliza en el filtrado de spam para correos electrónicos y análisis de opiniones, etc.

nuevas imágenes de los titulares. Para detectar las noticias falsas Este es un problema de clasificación más antiguo y se ocupa de grandes Barnaghi P [10] ha propuesto resolver los problemas de las noticias falsas utilizando dos fases. En la fase uno ni se detectan los usuarios falsos que ignoran las fake news ni pasan a la fase sucesora. Luego, averigüe si las noticias son confiables utilizando algoritmos de aprendizaje automático. El modelo propuesto ha encontrado al usuario falso en la primera fase mediante el uso del algoritmo Naïve Bayes con un 99% de precisión. En la segunda fase, las medidas de similitud con los algoritmos de aprendizaje automático encontraron las noticias falsas con un 95% de precisión. El trabajo futuro se realizará en sitios de redes sociales y en varios idiomas.

P (I / J) = (P (J / I) * P (I)) / P (J) Posterier = (Probabilidad * anterior) / Evidencia

P (I): Probabilidad previa de clase

P (J | I): probabilidad, la probabilidad del predictor dada la clase P (J): probabilidad previa del predictor

P (I | J): Probabilidad posterior fuera de la clase I dado el predictor J

3 TÉCNICAS DE PNL UTILIZADAS EN FALSAS NOTICIAS DETECCIÓN

4.2 Vecino más cercano

El algoritmo de vecino más cercano k es un algoritmo de aprendizaje automático supervisado fácil y simple que resuelve problema de clasificación y regresión. Consideremos algunos

El procesamiento del lenguaje natural ha venido de la extensión a problema de clasificación y regresión. Consideremos algunos informática e inteligencia artificial que son los puntos etiquetados y luego calcular la distancia entre cada interacción entre los humanos y las computadoras. NLP tiene apuntar y ordenar las distancias en orden decreciente y etiquetar el introducido para procesar la gran cantidad de datos en un primer punto significativo como "k". Las etiquetas K devuelven la media y la moda de la manera, modelos de regresión y clasificación. Este es un modelo sensible.

Collobert [11] introdujo el procesamientó del lenguaje natural a partir de Y los problemas de clasificación tienen valores discretos como salidas. K la arquitectura y los algoritmos de redes neuronales unificadas. Este más cercano es un modelo de detección de intrusiones y reconocimiento de patrones. 4.3 Máquina de vectores de soporte dice que cómo diferenciar palabras y oraciones que son

- hecho por el hombre. Las representaciones de vectores de palabras como: SVM es un algoritmo de aprendizaje supervisado. Comparando con otros **INCLINARSE**: Modelo de bolsa de palabras, las oraciones son algoritmos organizados, este es uno de los algoritmos que mejor se I. ajustan. SVM se utiliza en conjuntos múltiples. No funciona en la lingüística computacional de orden y contexto para encontrar las noticias falsas. Este algoritmo de la palabra ocurrencia. se utiliza para convertir los modelos de aprendizaje, que son solo para
 - TF-IDF: El término Frecuencia: uso específico del documento inverso. La normalización se realiza en los conjuntos de entrenamiento y II. la frecuencia pondera la palabra para mostrar los conjuntos de datos. Ha logrado un buen alcance en el conjunto entrenado. Apoye la importancia de esa palabra en una oración. vector machine es un algoritmo de clasificación ampliamente utilizado.
 - Guante: Primero construya la matriz de Co-ocurrencia y se usa principalmente para extraer una gran cantidad de datos y III. luego reduzca las dimensiones con la matriz por la cantidad de dimensión de la imagen. método de factorización.
 - IV. Word2VEC: Predice si la palabra dada está presente o nο.

4.4 Árboles de decisión

El árbol de decisiones es un modelo de aprendizaje supervisado y se representa como diagramas de flujo. Es un método no paramétrico eficiente que se puede utilizar para clasificación y regresión. En el árbol de decisión dividimos el conjunto de fuentes en subconjuntos basados en

prueba de valor de atributo. El proceso de división que se repite en cada modelo es el tamaño de los datos y los datos entrenados. El modelo de CNN se toma en un subconjunto de forma recursiva. El subconjunto en un nodo tiene la misma consideración en función del rendimiento y la velocidad. variable entonces podemos decir que la recursividad está completa. El final

El resultado es el árbol de decisión con nodos de decisión y nodos de hoja. 5.3 Redes neuronales recurrentes

Hay dos o más ramas para un nodo de decisión en las que el modelo RNN se entrena mediante retropropagación. Y esta es una El nodo hoja representa una clasificación o una decisión. Puede tipo de modelo ANN. Esta es también una red de avance manejar datos categóricos y numéricos. toma la entrada de bucles recurrentes. RNN realiza el análisis o

Tabla 1: Modelos de clasificación de aprendizaje automático PNL tabla de comparación de técnicas

Algoritmo Sra	Bayes ingenuo	SVM	Decisión ^{Árbol}	KNN
Métrica	ACC: 93,95% PREC: 95,98 % REC: 95,36% FM: 95,67%	ACC: 98,82 % PREC: 98,6 4% REC: 98,87 % FM: 98,23%	ACC: 91,93 % PREC: 95,5 1% REC: 92,96 % FM: 94,22%	ACC: 94,1% PREC: 89% REC: 86,2% FM: 81,5%
Parame ters	Precisión, Precisión, ^{Recordar,} Medida F	Precisión, Precisión, _{Recordar} , Medida F	Precisión, Precisión, _{Recordar,} Medida F	Precisión, Precisión, Recordar, Medida F
Ventaja ges	Fácilmente implementado , requiere menos entrenamiento período para probar los datos, rapido decisión haciendo	Eficiente para pequeña capacitación datos conjuntos SVM es Mejor modelo comparado a ANN	Fácil par explicar	a Fácil par a implementar
Disadva ntages	Pérdida de precisión	Toma mas duración a entrenar grande conjuntos de datos	Menos preciso	Computatio n costo es
Usos	Filtrado, Texto Clasificación	Intrusión detección, Mano escribiendo reconocimiento	Remoto sintiendo Planificación	Patrón reconocimiento, la medida documento semejanza

5 ALGORITMOS DE APRENDIZAJE PROFUNDO UTILIZADOS EN DETECCIÓN DE FALSAS NOTICIAS

5.1 Árboles de decisión

Se propusieron redes neuronales profundas para imitar los cerebros humanos para reconocer patrones. DNN es una red neuronal con redes, que contienen la capa de entrada, la capa de salida y una única capa oculta. Encuentra manipulaciones matemáticas para activar las relaciones de entrada y salida no lineales o lineales. Esta es una red de retroalimentación en la que los datos fluyen de la entrada a la salida sin retroceder. Estos están entrenados con propagación hacia atrás.

5.2 Redes neuronales convolucionales

Las redes neuronales convolucionales son una red de neuronas conectadas en capas que se utilizan para tomar las entradas y subcontratar la salida. CNN es un modelo de red de avance adecuado para el reconocimiento de objetos y el análisis de imágenes. La red neuronal se representa como etapas para realizar las funciones, hay tres etapas, como capa de convolución, capa de detector, capa de agrupación. El trabajo de la capa convolucional es construir mapa de características intrincado. La capa del detector es prominente

mapa de características intrincado. La capa del detector es prominente componentes no lineales de mapas de características. La capa de agrupación reduce**6. CONCLUSIÓN**

la información del predecesor y da la salida. CNN encuentra Debido al crecimiento de las tecnologías emergentes, la tasa de caracteres inactivos en el contenido de las noticias. El uso principal de este uso de las redes sociales se ha incrementado rápidamente. Hoy en día el

tipo de modelo ANN. Esta es también una red de avance que toma la entrada de bucles recurrentes. RNN realiza el análisis de datos de manera secuencial, como análisis de sentimientos, reconocimiento de voz y en alguna otra tarea. RNN es un modelo con memoria. Que toma las entradas del modelo anterior. Este modelo comprende el lenguaje humano y responde en consecuencia. Ejemplo de RNN es Siri de Apple y Alexa de Amazon. Esto no puede predecir el trabajo futuro basado en datos pasados. Recuerda la información pasada y utiliza los mismos parámetros para entradas adicionales o

5.4 Redes neuronales artificiales

capas ocultas para subcontratar la salida.

ANN es un algoritmo computacional. Las estructuras topológicas fueron imitadas con patrones complejos y no lineales. En las clasificaciones de imágenes de satélite se utilizan ANN. ANN es similar a la neurona humana. Eso envía señales eléctricas. Incluye una gran cantidad de unidades procesadoras conectadas que funcionan juntas. Este es un modelo de avance con capa de entrada, capa de salida y capas ocultas. El propósito de la capa de entrada es recibir entradas y comunicarse con la capa oculta. En consecuencia, la capa oculta se combina con la capa de entrada y envía una respuesta a la capa de salida. Alimentamos la red neuronal con algunas entradas y salidas para comparar la salida real con la salida ANN obtenida.

Tabla 1: Modelos de clasificación de aprendizaje profundo tabla de comparación

Algoritmos	CNN	RNN	ANA
Métrica	Precisión: 85% Precisión: 85% Recordar: 85% FM: 85%	Precisión: 76,9%	Precisión: 89,4% Precisión: 88% Recuerdo: 79,4% F-1: 84%
Ventajas	A una vez múltiple clasificaciones están hechos, localiza objetos en un imagen	No proceso límite, poderoso	Poder ACCESO entrenamiento múltiple algoritmos, necesitar menos capacitación
Desventajas	Computacional el costo es costoso, requieren mucho datos de entrenamiento	Escalada	Económico, sobreajuste
Usos	Imagen reconocimiento, tiempo serie previsión	Habla reconocimiento, Automático imagen etiquetador Máquina traducción	Imagen reconocimiento, Escrito a mano personaje reconocimiento, Facial reconocimiento

Las redes sociales se han convertido en las noticias diarias del mundo en lugar de los canales de noticias y los periódicos. La difusión de noticias falsas puede provocar una destrucción generalizada. Estamos difundiendo las noticias falsas a nuestro entorno mediante la confianza digna en las redes sociales, finalmente, que se está engañando demasiada confianza. En el futuro, no podremos distinguir entre noticias reales y noticias falsas, ya que la tasa de noticias falsas se ha convertido en un disfraz. En este artículo, hemos realizado una encuesta para detectar noticias falsas utilizando técnicas de procesamiento de lenguaje natural, técnicas de Deep Learning y algoritmos probados en algunos de los conjuntos de datos como LIAR, Buzzfeed, PolitiFact, Kaggle y McIntire y algoritmos de aprendizaje automático como Naïve-Bayes ', SVM, Decision Tree, AdaBoost, etc. La comparación de la precisión obtenida con vainilla y LSTM son menos precisas que GRU y CNN.

REFERENCIAS

- [1] Anuse, A. y Vyas, V. Complex Intell. Syst. (2016) 2: 221. https://doi.org/10.1007/s40747-016-0024-6.
- [2] Amol Agrawal. 2016. Detección de clickbait mediante aprendizaje profundo. En Tecnologías de Computación de Próxima Generación (NGCT), 2016 2da Conferencia Internacional sobre. IEEE. 268–272.
- [3] MEMORIA LARGA A CORTO PLAZO Computación neuronal 9 (8): 1735.
- [4] Jain N., Kumar A., Singh S., Singh C., Tripathi S. (2019)
 Detección de revisiones engañosas mediante técnicas de aprendizaje profundo. En: Métais E., Meziane F., Vadera S., Sugumaran V., Saraee M. (eds) Procesamiento del lenguaje natural y sistemas de información. NLDB 2019. Lecture Notes in Computer Science, vol 11608. Springer, Cham.
- [5] E. Van Der Walt y J. Eloff, "Uso del aprendizaje automático para detectar identidades falsas: Bots vs humanos", en *Acceso IEEE*, vol. 6, págs.6540-6549, 2018.
- [6] ML Della Vedova, E. Tacchini, S. Moret, G. Ballarin, M. DiPierro y L. de Alfaro, "Detección automática de noticias falsas en línea que combinan contenido y señales sociales", 2018 22a Conferencia de la Asociación de Innovaciones Abiertas (FRUCT), Jyvaskyla, 2018, págs. 272-279.
- [7] Wu, L. y Liu, H. (2018). Seguimiento de huellas de noticias falsas. Actas de la Undécima ACM International Conferencia sobre búsqueda web y minería de datos WSDM '18. doi: 10.1145 / 3159652.3159677.
- [8] Ahmed H., Traore I., Saad S. (2017) Detección de noticias falsas en línea mediante análisis N-Gram y técnicas de aprendizaje automático. En: Traore I., Woungang I., Awad A. (eds) Sistemas inteligentes, seguros y confiables en entornos distribuidos y en la nube. ISDDC 2017. Lecture Notes in Computer Science, vol 10618. Springer, Cham.
- [9] Elazab, Ahmed. (2018). Detección de noticias fraudulentas para redes sociales online.
- [10] Barnaghi P., Kontonatsios G., Bessis N., Korkontzelos Y. (2019) Extracción de aspectos de revisiones que utilizan redes neuronales convolucionales e incorporaciones. Procesamiento del lenguaje natural y sistemas de información. NLDB 2019. Lecture Notes in Computer Science, vol 11608. Springer.