Sistemas expertos con aplicaciones 184 (2021) 115412

Listas de contenidos disponibles en ScienceDirect

Sistemas Expertos con Aplicaciones

revista Página de inicio: www.elsevier.com/locate/eswa

AMFB: agrupación bilineal factorizada multimodal basada en la atención para la detección multimodal de noticias falsas

Rina Kumari *, Asif Ekbal

Departamento de Ingeniería y Ciencias de la Computación, Instituto Indio de Tecnología de Patna, India

INFORMACIÓN DEL ARTÍCULO

Palabras clave:
Detección multimodal de noticias falsas
Aprendizaje profundo
Mecanismo de atención
Fusión de características multimodales
Combinación bilineal factorizada multimodal

ABSTRACTO

Las noticias falsas son la información o historias que se crean intencionalmente para engañar o engañar a los lectores. En los últimos tiempos, la detección de noticias falsas ha atraído la atención de investigadores y profesionales debido a sus múltiples beneficios, incluida la adopción de medidas preventivas para abordar la difusión de información errónea que, de otro modo, podría perturbar el tejido social. Las redes sociales en los últimos tiempos están muy cargadas de noticias e información multimedia. Las personas prefieren la lectura de noticias en línea y lo encuentran más informativo y conveniente si tienen acceso a contenido multimedia en forma de texto, imágenes, audio y videos. En los primeros estudios, los investigadores han propuesto varios mecanismos de detección de noticias falsas que utilizan principalmente las características textuales y no son adecuados para aprender la representación compartida multimodal (textual + visual).

Para superar estas limitaciones, en este artículo, proponemos un marco de detección de noticias falsas multimodal con una fusión de características multimodal apropiada que aprovecha la información del texto y la imagen y trata de maximizar la correlación entre ellos para obtener la representación compartida multimodal eficiente. Demostramos empíricamente que el texto, cuando se combina con la imagen, puede mejorar el rendimiento del modelo. El modelo detecta la publicación una vez que se introduce en la red en una etapa temprana. En la etapa inicial de la introducción de una publicación de noticias en la red, el modelo toma el texto y la imagen de la publicación como entrada y decide si es falso o genuino. Dado que este modelo solo analiza contenidos de noticias, no requiere ninguna información previa sobre el usuario y detalles de la red. Este marco tiene cuatro submódulos diferentes*verbigracia.* **Memoria bidireccional apilada a corto plazo basada en la atención (ABS-BiLSTM)** para la representación de características textuales, **Red neuronal convolucional multinivel basada en la atención: red neuronal recurrente (ABM-CNN-RNN)** para la extracción de características visuales, **agrupación bilineal factorizada multimodal (MFB)** para la fusión de funciones y finalmente **Perceptrón multicapa (MLP)** para la clasificación. Realizamos experimentos en dos conjuntos de datos disponibles públicamente,*verbigracia.* Twitter y Weibo. Los resultados de la evaluación muestran la eficacia de nuestro enfoque propuesto que funciona significativamente mejor en comparación con los modelos de última generación. Demuestra que supera el estado actual de la técnica en aproximadamente 10 puntos para el conjunto de datos de Twitter. Por el contrario, el conjunto de datos de Weibo logra un mejor rendimiento general con puntuaciones F1 equilibradas entre clases reales y falsas. Además, la complejidad de nuestro modelo propuesto es significativamente menor que el estado de la técnica.

1. Introducción

El contenido de noticias engañosas y falsas en las redes sociales es uno de los desafíos considerables en nuestra sociedad. Las noticias falsas se pueden definir como la información creada intencionalmente mediante la manipulación de texto, imágenes, audios o videos. Algunas plataformas de redes sociales populares, como Twitter, Facebook y blogs, desempeñan un papel inevitable en la rápida difusión de noticias. Dado que el número de lectores de noticias en línea aumenta continuamente, algunas personas aprovechan la difusión de información falsa para engañarlos. Esto puede generar un impacto negativo en la sociedad e incluso manipular eventos públicos importantes. El presidencial de Estados Unidos La elección de 2016 da una mejor realización de la misma. Durante esta elección, se generaron noticias falsas para apoyar a cualquiera de los dos candidatos. Mucha gente había creído y también compartido más de 37 millones de veces en Facebook (Wang y col., 2018).

Figura 1 describe algunos ejemplos de noticias falsas multimodales del conjunto de datos de Twitter, que usamos para la evaluación (los detalles se encuentran en las secciones siguientes). Cada tweet contiene un fragmento de texto asociado con una imagen. La imagen del primer tweet ha sido fotografiada, pero el texto es real porque el eclipse solar fue un evento real, pero la vista fue diferente. El segundo tweet muestra la imagen real, pero esta es una imagen de un

* Autor correspondiente.
Correos electrónicos: rina_1921cs13@iitp.ac.in (R. Kumari), asif@iitp.ac.in (A. Ekbal).
URL: http://www.iitp.ac.in/~asif/ (A. Ekbal).

https://doi.org/10.1016/j.eswa.2021.115412
Recibido el 12 de agosto de 2020; Recibido en forma revisada el 11 de marzo de 2021; Aceptado el 9 de junio de 2021
On-line el 29 de junio de 2021

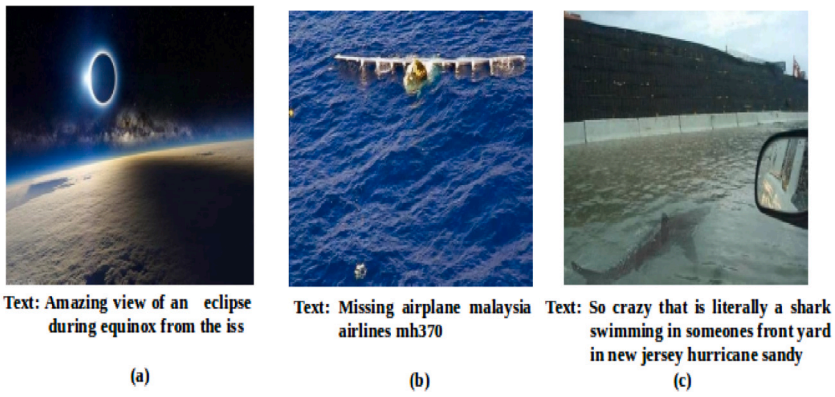

Figura 1. Ejemplo de noticias falsas del conjunto de datos de Twitter.

accidente de avión en Sicilia. En la parte derecha de la imagen, se ha creado un tiburón artificial que no existía durante el huracán Sandy. Estos tweets son potencialmente falsos y su principal propósito de difusión es engañar a la población. Es una tarea muy desafiante detectar este tipo de información como falsa o real.

Definición del problema: En nuestro trabajo actual, abordamos el problema de la detección de noticias falsas aprovechando la información de múltiples fuentes, como texto e imágenes. En particular, nos enfocamos en utilizar y fusionar contenido textual y visual para detectar una publicación multimedia como falsa o real. Formulamos este problema de la siguiente manera.

Supongamos que se nos da un conjunto de 'm' número de publicaciones de noticias multimedia $P = (_1, _2, \ldots,)$. *Cada publicación*( ) *contiene textos* $T = (_1, _2, \ldots,)$, *imágenes correspondientes* $I = (_1, _2, \ldots,)$ *y etiquetas* $Y = (_1, _2, \ldots,)$. *Contenido del texto* ( ) *de la publicación de noticias se compone de una sola oración o grupo de oraciones. Etiqueta*( ) *se da para cada publicación* ( ). *Un clasificador 'C' debe ser entrenado con la ayuda de un conjunto dado de publicaciones de noticias. El clasificador toma todo el contenido del texto junto con la imagen asociada de una noticia correo*( ) *como entrada para clasificar la publicación dada como falsa* (= 0) *o real* (= 1), *es decir* (= (,)). *Aquí,* , *y son la etiqueta predicha, la instancia de texto y la imagen correspondiente de la publicación de noticias, respectivamente* $yk \in [1 \ldots]$.

Motivación y contribución: Muchos métodos de aprendizaje tradicionales y, en los últimos tiempos, los modelos de aprendizaje profundo se han puesto a disposición para la detección de noticias falsas. Sin embargo, la atención se ha centrado principalmente en el texto, por ejemplo,Castillo, Mendoza y Poblete (2011). Ejemplos enFigura 1 mostrar que el contenido de noticias falsas no se puede identificar correctamente utilizando una única modalidad, es decir, información de modalidad de texto o imagen. La identificación podría ser posible si aprovechamos la información de diversas fuentes (es decir, texto e imagen), lo que da lugar al concepto de detección multimodal de noticias falsas.

Algunas de las obras anteriores, como Khattar, Goud, Gupta y Varma (2019), Wang y col. (2018) han tratado de conocer las características compartidas entre todos los eventos y la correlación entre texto e imagen. Estos mecanismos otorgan la misma importancia a todas las partes del texto y la imagen, lo que puede no ser un enfoque correcto para fusionar la información. EnSinghal, Shah, Chakraborty, Kumaraguru y Satoh (2019) el autor utiliza Representaciones de codificador bidireccional de Transformers (BERT) (Devlin, Chang, Lee y Toutanova, 2018) y modelo VGG19 previamente entrenado (Shaha y Pawar, 2018) para la extracción de características textuales y visuales, respectivamente. Después de la extracción, concatenaron ambas características que no proporcionan una buena correlación entre el texto y la imagen. Es posible que no detecte noticias engañosas (las noticias que se muestran en el segundo ejemplo deFigura 1) correctamente.

En la literatura, los autores utilizaron principalmente la red neuronal convolucional (CNN) previamente entrenada como VGG19 (Shaha y Pawar, 2018) para la extracción de características visuales. El VGG19 se entrenó en el conjunto de datos ImageNet (Deng, Dong, Socher, Li, Li y Fei-Fei, 2009), que es más adecuado para la detección de objetos. Como el conjunto de datos de ImageNet es un conjunto de datos de dominio general, es posible que VGG19 no capture la semántica específica del dominio. características de las imágenes de noticias falsas debido a la falta de información relevante para la tarea. Por lo tanto, la extracción de las características inherentes de las imágenes de noticias falsas es una tarea muy desafiante. Aparte de la extracción de características, la fusión de características también es una tarea muy importante. Los autores acaban de concatenar características textuales y visuales para obtener una representación conjunta de características multimodales en las obras existentes. En base a esto, han clasificado las publicaciones de noticias como falsas o reales.

Inspirados por la falta de una buena extracción de características y mecanismos efectivos de fusión de características, proponemos un nuevo marco basado en una red neuronal profunda para la detección multimodal de noticias falsas en este trabajo. Este marco se denomina "**A**basado en la atención **METRO**ultimodal **F**actorizado **B**agrupación ilineal (AMFB) ".

Resumimos las contribuciones de este trabajo de la siguiente manera:

(I). Proponemos un LSTM bidireccional apilado basado en la atención (ABS-BiLSTM) red que captura información textual en diferentes niveles.

(ii). Proponemos una Neural Network – Red neuronal recurrente (ABM-CNN – RNN) para la extracción de características visuales que extrae características inherentes de la imagen.

(iii). Combinamos las representaciones de características textuales y visuales. utilizando el módulo MFB y pasarlo a través de un modelo de perceptrón multicapa (MLP) con dos capas ocultas y una capa de salida con una función de activación sigmoidea para la detección de noticias falsas.

(iv). Realizamos extensos experimentos en dos conjuntos de datos de referencia para validar el desempeño del modelo propuesto. La evaluación muestra que el enfoque propuesto alcanza un nuevo desempeño de vanguardia.

El resto de este documento está organizado de la siguiente manera. En la sección2, presentamos una breve reseña de los trabajos relacionados. Sección3 explica los objetivos de la investigación. Sección4 describe la metodología en detalle. En la sección5, primero describimos los conjuntos de datos utilizados para nuestros experimentos y demostramos los experimentos realizados, seguidos de un análisis detallado. Sección6 concluye nuestro trabajo junto con algunas hojas de ruta para la dirección futura.

2. Trabajo relacionado

Un artículo de noticias contiene varios aspectos como fuente, autor, titulares pegadizos, estilos de redacción, imágenes, vídeos, etc. Cualquier cambio realizado en estos aspectos genera noticias engañosas o falsas. La tarea de detección de noticias falsas es algo similar a muchas otras tareas, como la detección de rumores (Jin, Cao, Jiang y Zhang, 2014; Zhang, Fang, Qian y Xu, 2019), detección de spam (Liu, Pang y col., 2019; Shen y col., 2017) y detección de sátiras (Rubin, Conroy, Chen y Cornwell, 2016). La detección de rumores es muy similar a la detección de noticias falsas porque las noticias falsas son un tipo de rumor falso.

El método propuesto en Alkhodair, Ding, Fung y Liu (2020) detects rumores de noticias de última hora en lugar de rumores de larga duración. La teoría detrás de este mecanismo es**los rumores de noticias de última hora se difundieron rápidamente**

en comparación con los rumores de larga duración debido a la menor conciencia y los algoritmos de verificación automática de hechos que requieren menos tiempo. Demuestra un enfoque que aprende simultáneamente word2vec (Iglesia, 2017) incrustaciones de palabras y redes neuronales recurrentes con múltiples objetivos para identificar automáticamente los rumores de noticias de última hora. Word2vec proporciona vectores de palabras independientes del contexto, y RNN simple solo se encarga de la representación secuencial de las palabras en el texto, pero no extrae la información vital del texto. Por lo tanto, el rendimiento de este marco se puede mejorar utilizando algunas incrustaciones de palabras contextuales como BERT y un mecanismo de atención sobre RNN durante la implementación.

El trabajo propuesto en Liu, Jin y col. (2019) intenta capturar la dinámica cambios de contenidos de noticias, difusores de noticias y estructuras de difusión. Los autores de este artículo han implementado un modelo de detección temprana de rumores basado en la memoria larga a corto plazo (LSTM) para identificar los rumores en la etapa inicial. Es una idea novedosa para capturar las diferencias dinámicas entre las estructuras de difusión y los difusores de rumores y no rumores. Sin embargo, si el mismo mensaje se escribe en una secuencia de palabras diferente, el modelo falla. Más específicamente, da predicciones incorrectas en el caso de parafrasear. En otro trabajo Zubiaga y col., 2018), los autores han demostrado que el rendimiento del clasificador secuencial aumenta si explota las características del discurso extraídas de las interacciones de las redes sociales. Este artículo también muestra que el LSTM supera al clasificador secuencial mientras utiliza un conjunto reducido de características. Dado que este es el mecanismo basado en funciones, la atención sobre LSTM puede mejorar el rendimiento seleccionando información importante del conjunto de funciones seleccionado. Los investigadores han desarrollado varios enfoques para detectar noticias falsas y dejar de difundirlas en las redes sociales. A continuación presentamos una revisión de los trabajos existentes en dos amplias categorías: (i). Detección de noticias falsas basada en una única modalidad, es decir, unimodal; y (ii). Detección de noticias falsas basada en multimodalidad.

2.1. Detección de noticias falsas basada en modalidad única

La detección de noticias falsas se ha llevado a cabo predominantemente solo para texto, pero los autores también se han centrado en la información visual y contextual en estudios recientes. Las características textuales son generalmente las características semánticas o estadísticas extraídas del texto. Los autores en Faustini y Covões (2020) propuso un mecanismo para detectar noticias falsas utilizando solo características de texto. Genera características textuales independientes de la plataforma de origen e independientes del idioma y aplica Naive Bayes (NB) (Rish y col., 2001), Bosque aleatorio (RF) (Gilda, 2017), K-Vecino más cercano (KNN) (Zhang, 2016) y Support Vector Machine (SVM) (Huang y col., 2018) para la clasificación de noticias falsas. Este mecanismo realiza una especie de ingeniería de funciones que no puede ser una solución más rápida y automática para la detección de noticias falsas. Este trabajo se puede mejorar aún más si aprende automáticamente el lenguaje y la representación de funciones independientes de la plataforma y realiza la detección de noticias falsas utilizando algoritmos de aprendizaje profundo de extremo a extremo. El trabajo reportado enCastillo y col. (2011) propuso una evaluación automática de la credibilidad utilizando técnicas de clasificación como SVM y Decision Tree (DT) (Canción y Ying, 2015) para identificar la credibilidad de los tweets en función de las características del usuario y del tweet, que son principalmente de tipo estadístico y semántico. En este trabajo, el rendimiento se puede mejorar utilizando un modelo de extracción automática de características porque las características estadísticas no capturan los cambios dinámicos en la propagación a lo largo del tiempo.

En Gravanis, Vakali, Diamantaras y Karadais (2019), los autores han introducido un modelo de conjunto basado en aprendizaje automático que utiliza funciones basadas en contenido y algoritmos de aprendizaje automático para la detección de noticias falsas. Los algoritmos de aprendizaje automático funcionan bien, pero es un desafío obtener características textuales hechas a mano para los modelos tradicionales de detección de noticias falsas basados en el aprendizaje automático debido a la falta de conocimiento y experiencia en el dominio. (Kwon, Cha, Jung, Chen y Wang, 2013; Rashkin, Choi, Jang, Volkova y Choi, 2017) identificó los rumores a partir de los tres aspectos de la difusión: estructural, temporal y lingüística. El autor ha demostrado las diferencias lingüísticas y estructurales clave durante la difusión de publicaciones de rumor y no rumor.

Los métodos propuestos extraen las características lingüísticas del texto engañoso. Extraen las características analizando en qué se diferencia el patrón de lenguaje de las noticias reales de los engaños, la sátira y la propaganda. Con base en estas características lingüísticas, los modelos clasifican la publicación de noticias como verdadera o falsa. Las características lingüísticas dependen en gran medida del conocimiento del dominio y eventos específicos; los patrones lingüísticos aún no se comprenden bien. Por lo tanto, los modelos se pueden fortalecer aún más mejorando las características lingüísticas extraídas.

Autores de Potthast, Kiesel, Reinartz, Bevendorff y Stein (2018) han informado de un mecanismo que analiza el estilo de las noticias falsas y la información hiperpartidista (extremadamente sesgada). Muestran cómo las noticias convencionales y las hiperpartidistas se pueden distinguir mediante el análisis de estilo. Los estilos de redacción de las mismas noticias dificultan la tarea de detección de noticias falsas y hacen predicciones erróneas.

Además de las características textuales y lingüísticas, el contexto social también proporciona evidencia útil para la detección de noticias falsas. Recientemente, enShu, Wang y col. (2019), los autores han descrito cómo se utilizan los contextos sociales en la detección de noticias falsas. Este artículo propuso un marco para la incorporación de tres relaciones llamado TriFN que modela simultáneamente las interacciones entre el usuario y las noticias y las relaciones entre el editor y las noticias para la clasificación de noticias falsas.Shu, Cui y col. (2019) explica por qué una noticia o publicación en particular se detecta como falsa. Los autores han desarrollado una subred de atención conjunta de oraciones y comentarios que explota los contenidos de las noticias junto con los comentarios de los usuarios para aprender y capturar de manera conjunta oraciones y comentarios de los usuarios para la detección explicable de noticias falsas. Dado que las características del contexto social no están estructuradas, son muy ruidosas y su recopilación requiere mucha mano de obra, no puede proporcionar información suficiente y relevante para los eventos recién surgidos. El trabajo presentado enShu, Wang y col. (2019) y Shu, Cui y col. (2019) puede mejorarse si se refuerzan las características del contexto social y sus mecanismos de selección.

Dado que todos los modelos discutidos anteriormente están basados en características, se requiere conocimiento del dominio para obtener un buen conjunto de características, lo que requiere mucho tiempo y muchos esfuerzos. Por el contrario, los modelos basados en el aprendizaje profundo como (Huang y Chen, 2020; Ma et al., 2016) puede extraer las características automáticamente de los datos y, por lo tanto, no necesita ningún conocimiento profundo. EnHuang y Chen (2020), los autores han introducido un mecanismo basado en conjuntos que combina diferentes modelos de aprendizaje profundo para la detección de noticias falsas. EnRuchansky, Seo y Liu (2017), el autor ha propuesto un modelo basado en redes neuronales recurrentes (RNN) que intenta averiguar los patrones de las actividades del usuario en una publicación determinada y decide si la publicación es falsa o real en función de las actividades del usuario. Aunque los modelos de aprendizaje profundo extraen las características automáticamente, a menudo pueden contener ruido y características irrelevantes que degradan el rendimiento del modelo.

Recientemente, en Wu, Rao, Nazir y Jin (2020), los autores han explicado cómo las características extraídas a través del aprendizaje profundo sufren de muchas características ruidosas e irrelevantes que reducen el rendimiento de los enfoques. Han propuesto un modelo novedoso basado en redes neuronales Adversarial con el objetivo de reducir las características irrelevantes y redundantes de las características extraídas para medir la credibilidad de la información. Aunque el entrenamiento adversario del modelo ofrece mejores características, es difícil de entrenar. Una vez más, el modelo es muy complejo de ejecutar con recursos limitados, y también lleva mucho tiempo generar las características relevantes y silenciosas. Autores enKarimi, Roy, Saba-Sadiya y Tang (2018) introdujo un enfoque que combina información de múltiples fuentes y propuso un marco denominado Detección de noticias falsas de múltiples fuentes y clases múltiples (MMFD). También diferencia entre los diferentes grados de falsedad. Finalmente, este marco combina grados automatizados de falsedad, extracción automatizada de características y fusión de múltiples fuentes en un modelo interpretable y coherente para la detección de noticias falsas. Este trabajo solo considera el texto de la noticia de las diferentes fuentes y perspectivas, pero no se preocupa por los comentarios y respuestas. Dado que la difusión de noticias falsas también depende de la evidencia proporcionada en los comentarios, el trabajo se puede mejorar aún más si incluye los comentarios en esa publicación de noticias.

Existe una literatura muy limitada que utiliza características visuales para verificar publicaciones multimedia. Jin, Cao, Zhang, Zhou y Tian (2016) han demostrado la importancia de una función de imagen para la verificación automática de noticias falsas en las redes sociales. Ha demostrado que los eventos de noticias reales y falsas contienen diferentes patrones de distribución de imágenes. (ping Tian y col., 2013) explicó la extracción de características de la imagen y el mecanismo de representación. Aquí, el autor analizó el rendimiento de los modelos de detección de noticias falsas después de fusionar características locales y globales de las imágenes. El trabajo presentado enJin y col. (2016), ping Tian y col. (2013) se basan completamente en las características de la imagen. Sin embargo, estas características de la imagen todavía están hechas a mano y difícilmente pueden representar distribuciones complejas de contenido visual. Los investigadores deben trabajar en la extracción automática de características de imágenes mediante la implementación de algunos mecanismos basados en el aprendizaje profundo.

2.2. Detección de noticias falsas basada en multimodalidad

En los últimos tiempos, el análisis de información multimodal ha atraído la atención de investigadores y profesionales para resolver varios problemas prácticos como el análisis de sentimientos (Ghosal y col., 2018), análisis de emociones (Chauhan, Akhtar y col., 2019), subtítulos de imágenes (Karpathy y Fei-Fei, 2015), respuesta visual a preguntas (Antol y col., 2015), y también detección de noticias falsas (Jin, Cao, Guo, Zhang y Luo, 2017) etc. En Jin y col. (2017), se ha elaborado un modelo basado en el aprendizaje profundo planteados para extraer características multimodales junto con características del contexto social y combinarlas mediante un mecanismo de atención. Por lo tanto, el modelo extrae características específicas de eventos y no se puede generalizar para detectar noticias falsas sobre eventos recién llegados. El rendimiento de este trabajo se puede mejorar combinando un mecanismo de extracción de características independientes de eventos y dominios con el modelo propuesto. Para superar la limitación encontrada enJin y col. (2017), Wang y col. (2018) propuso un modelo basado en el aprendizaje profundo denominado Event Adversarial Neural Network (EANN). El modelo genera representaciones de características invariantes de eventos con la ayuda de una red de adversarios (Goodfellow y col., 2014). Este modelo no es capaz de aprender la representación compartida de publicaciones multimodales. Para evitar este problema del aprendizaje de la representación,Khattar y col. (2019) introdujo el codificador automático variacional multimodal (MVAE) para la detección de noticias falsas. Una subtarea adicional (discriminador de eventos en EANN (Wang y col., 2018) y parte del decodificador del autoencoder variacional en MVAE (Khattar y col., 2019)) se han introducido para obtener los resultados. En estos modelos, los resultados dependen en gran medida de la subtarea, cuya ausencia degrada el rendimiento del modelo. Encontrar representaciones multimodales compartidas en ambos (EANN y VAE) estos trabajos es otra limitación. Los marcos de detección de noticias falsas presentados en estos dos artículos concatenan las representaciones de características de texto e imagen y, por lo tanto, obtuvieron características multimodales que pueden no introducir la interacción y alineación adecuadas entre características textuales y visuales. El autor enSinghal y col. (2019) ha introducido un marco multimodal (Spotfake) que detecta noticias falsas sin realizar ninguna subtarea. En este trabajo, el autor ha resuelto las primeras limitaciones encontradas en EANN y VAE, pero continúa con otra limitación de obtener mejores representaciones compartidas.

Como se describe en la literatura anterior, en cualquier marco de aprendizaje multimodal, uno de los temas cruciales es investigar las técnicas de fusión apropiadas para combinar de manera efectiva la información de múltiples fuentes. El trabajo presentado enFukui, Park, Yang, Rohrbach, Darrell y Rohrbach (2016) introdujo la agrupación bilineal compacta (MCB) multimodal que genera una representación de características conjuntas de muy alta dimensión basada en el producto externo de dos vectores de características. Para superar el problema de la alta dimensionalidad (Kim y col., 2016) presentó el agrupamiento bilineal de rango bajo multimodal (MLB). Este mecanismo realiza la multiplicación por elementos (operación del producto Hadamard) de dos vectores de características para obtener la representación de características conjuntas en el espacio compartido. Recientemente,Chauhan, Firdaus y col. (2019) y Yu, Yu, Fan y Tao (2017) utilizó el módulo MFB para obtener la representación fusionada de funciones de texto e imagen para la generación de lenguaje natural (NLG) (Reiter y Dale, 1997) en el dominio de la moda.

Los trabajos anteriores con modalidad única utilizaban principalmente características artesanales para la extracción de características textuales. Algunos trabajos unimodales y multimodales existentes utilizaron RNN para la extracción de características textuales. Para la extracción de características visuales en noticias falsas multimodales, los autores de detección han utilizado la red VGG19 previamente entrenada. La extracción de características hecha a mano necesita experiencia en el dominio y requiere mucho tiempo. Es posible que RNN no extraiga las características semánticas de alto nivel, y una red VGG19 previamente entrenada, entrenada en el conjunto de datos ImageNet de dominio general, puede que no extraiga características de imagen específicas del dominio. En trabajos de investigación anteriores, los autores tampoco han diseñado ningún mecanismo de fusión de características multimodal. Proponemos una arquitectura CNN-RNN multinivel basada en la atención para la extracción de características visuales, y Bi-LSTM apilado basado en la atención para la extracción de características textuales para superar estas limitaciones en el foco de las deficiencias anteriores. Después de eso, combinamos ambas funciones utilizando el mecanismo MFB y pasamos la representación unificada a través de un perceptrón multicapa (MLP) para clasificar las publicaciones de noticias multimedia como reales o falsas. La evaluación de dos conjuntos de datos de referencia muestra que nuestro enfoque propuesto alcanza un rendimiento de vanguardia, posiblemente debido a una mejor extracción de características y un futuro mecanismo de fusión.

3. Objetivo de la investigación

En esta sección se presentan los objetivos específicos para la detección multimodal de noticias falsas. El objetivo principal de nuestra investigación es analizar el contenido de una publicación de noticias difundida en las redes sociales y detectar si es falsa o real. Nuestro trabajo actual presenta un análisis de hasta qué punto la fusión de diferentes modalidades (por ejemplo, texto, imagen) de una publicación de noticias mejora el rendimiento de la clasificación. Establecemos los siguientes cuatro objetivos principales de investigación:

RO 1. *Extracción de características mejores y relevantes del contenido de texto e imagen de una publicación de noticias multimedia.*

Nuestro primer objetivo de investigación tiene como objetivo extraer características útiles y relevantes del contenido de las noticias. Dado que consideramos diferentes modalidades, nos centramos en la extracción de características tanto de los textos como del contenido de imágenes de una publicación de noticias. Hacemos esto utilizando algunos marcos basados en aprendizaje profundo.

RO 2. *Proponga un mecanismo de fusión de funciones para maximizar la correlación entre las funciones de texto e imagen.*

Después de la extracción de características, otro objetivo es fusionar las características extraídas de texto e imagen para maximizar la correlación entre ellas y generar una mejor representación compartida. Logramos esto utilizando un mecanismo de agrupación bilineal factorizada multimodal (MFB).

RO 3. *Diseñe un marco de detección de noticias falsas multimodal novedoso que clasifique las publicaciones de noticias con alta precisión.*

El objetivo principal de este trabajo de investigación es diseñar un marco novedoso de detección de noticias falsas que aproveche la información del contenido de texto e imagen de las noticias, fusione esa información y decida si esta noticia es falsa o real. Hacemos esto implementando Multilayer Perceptron (MLP) sobre la representación compartida.

RO 4. *Evalúe la coherencia del modelo propuesto en diferentes conjuntos de datos.*

Nuestro objetivo final es construir un detector de noticias falsas multimodal que se generalice a los diferentes conjuntos de datos. Para lograr esto, evaluamos nuestro modelo propuesto en dos conjuntos de datos diferentes. Realizamos un análisis detallado para mostrar la efectividad, fortaleza y debilidades del modelo propuesto de detección de noticias falsas.

4. Metodología

Esta sección describe una intuición de los fundamentos teóricos y muestra cómo reformular el problema para permitir un cálculo preciso, eficiente y rápido de la credibilidad de las noticias. Hallazgos (mostrados enFigura 1 y discutido en el párrafo anterior) y el papel de múltiples modalidades en la detección de noticias falsas ofrecen una base teórica para un examen empírico adicional de las noticias falsas. Este trabajo sienta las bases

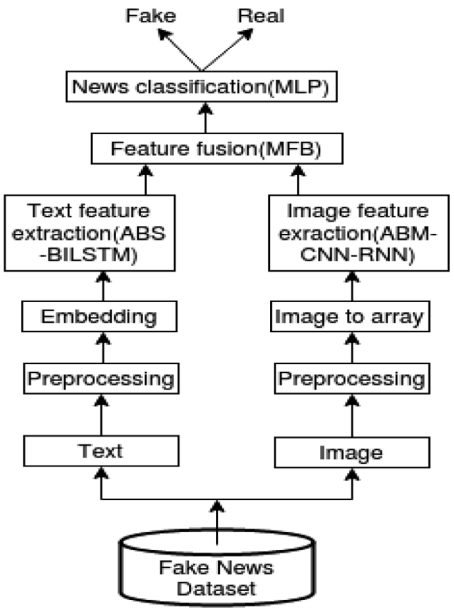

Figura 2. Diagrama de proceso del marco de detección de noticias falsas multimodal propuesto.

para crear un proceso repetible de un extremo a otro para detectar la difusión de noticias falsas multimodales en las redes sociales. Estos esfuerzos para caracterizar las publicaciones de noticias multimedia como falsas o reales ayudarán a los investigadores, periodistas, científicos y personas en general con la desinformación. Más específicamente, el documento presenta tres fundamentos esenciales para la detección de noticias falsas,*verbigracia*.(I). Extracción de características multimodal: en este paso, extraemos las características textuales y visuales de publicaciones de noticias multimedia; (ii). Fusión de características multimodales: en este paso, combinamos las características textuales y visuales extraídas para obtener una única representación compartida (iii). Detección multimodal de noticias falsas: En este último paso, clasificamos la publicación de noticias utilizando la representación compartida. Una vista de alto nivel de nuestra técnica se describe enFig. 3. Discutimos todos los fundamentos teóricos en detalle en esta sección.

Figura 2 muestra el diagrama de proceso general del marco de detección de noticias falsas multimodal propuesto. Con la ayuda de este diagrama, explicamos cómo se implementa la metodología propuesta desde el conjunto de datos sin procesar hasta la decisión final. Discutimos estas etapas una por una en detalle.

Conjunto de datos de noticias falsas: Inicialmente, tomamos los conjuntos de datos de noticias falsas multimodales disponibles públicamente. El conjunto de datos incluye id, fragmento de texto, imagen asociada, características de contexto y etiquetas de verdad fundamental.

Selección de características: Nuestro modelo actual se ocupa de la detección multimodal de noticias falsas, y consideramos un fragmento de texto, una imagen y los atributos de etiqueta de verdad del terreno correspondientes del conjunto de datos. También se utilizan atributos de texto e imagen extraídos de la red neuronal.

Preprocesamiento: En este paso preprocesamos el texto y la imagen. Inicialmente, eliminamos las instancias que solo contienen texto o imagen para preparar un conjunto de datos multimodal completo. Además, para los datos de texto, tokenizamos las oraciones, eliminamos los signos de puntuación y dejamos de palabras. Hacemos todas las imágenes de igual tamaño y convertimos las imágenes en un Matriz tridimensional.

Incrustación: Nosotros usar pre-entrenado texto rápido incrustación de 300 dimensiones para encontrar la palabra vectores.

Extracción de características: Después de encontrar incrustaciones de palabras de texto, implementamos **ABS-BiLSTM** para la extracción de características textuales. Para la extracción de características visuales, implementamos**ABM-CNN – RNN**.

Fusión de funciones: Después de obtener las representaciones de características textuales y visuales, implementamos el mecanismo de fusión de características para obtener la representación compartida. Mostramos una descripción detallada del mecanismo de fusión de características en la sección de metodología de este trabajo de investigación.

Clasificación de noticias: Una vez obtenida la representación combinada, la etapa final es la clasificación de la noticia. Implementamos el Perceptrón Multicapa (MLP) para la clasificación de noticias falsas y volvemos a poner la descripción en el apartado de metodología. Esta última etapa decide si la información es falsa o real. Extendemos este diagrama de proceso para diseñar y explicar nuestro modelo propuesto, que discutiremos más adelante en esta sección.

En este artículo, diseñamos una técnica basada en el aprendizaje profundo para la detección multimodal de noticias falsas. Utiliza información tanto textual como visual para decidir si una publicación de noticias multimedia es falsa o real. Definimos una instancia de publicación = *{TELEVISOR}* como una tupla que representa dos modalidades diferentes de una publicación de noticias, donde *T* y *V* representar textual y contenido visual, respectivamente. El modelo propuesto extrae textuales característica ( ) y característica visual ( ) para la instancia dada . La arquitectura general del modelo propuesto se describe en Fig. 3. Eso incluye cuatro componentes: (a). Bi-LSTM apilado basado en la atención (ABS-BiLSTM); (B). CNN – RNN multinivel basado en la atención (ABM-CNN – RNN); (C). Agrupación bilineal factorizada multimodal (MFB) y (d). Perceptrón multicapa (MLP).

Analizamos cada uno de estos módulos a continuación:

4.1. BiLSTM apilado basado en atención (ABS-BiLSTM)

Las diferentes modalidades producen diferentes aspectos de una publicación de noticias multimedia. Esta parte de la arquitectura propuesta extrae las características textuales de las publicaciones de noticias multimedia. Captura las mejores representaciones de características contextuales y semánticas de las palabras. El apilamiento de dos capas BiLSTM forma esta subred. Damos la salida de la primera capa BiLSTM como entrada a la segunda capa BiLSTM. Aquí, la segunda capa BiLSTM extrae características más complejas y algunos patrones diferentes sobre las características extraídas. Matemáticamente, lo formulamos como:

Dado una publicación de noticias , se emplean dos BiLSTM apilados para codificar cada palabra en los vectores ocultos h. Las palabras están representadas por 300-incrustación dimensional. Aquí, es el h publicación de noticias, , es el h palabra de h publicación de noticias, , ∈ (1,....,), ∈ (1,....,) y ∈ [1, 2].

$$\begin{aligned} \overrightarrow{h_{??}} &= \overrightarrow{(}, h_{??}\overrightarrow{]} \\ \overleftarrow{h_{??}} &= \overleftarrow{(}, \overleftarrow{h_{??}}\overleftarrow{]} \\ h, &= [\overrightarrow{h_{??}}, \overleftarrow{h_{??}}] \end{aligned} \quad (1)$$

Todas las palabras no contribuyen por igual a la representación del texto significativo. Aplicamos un mecanismo de atención sobre el estado oculto obtenido de la segunda capa BiLSTM para extraer palabras importantes y agregar la representación de esas palabras informativas para formar el vector de representación de texto final, similar aYang y col. (2016). Nosotros Formule este mecanismo de atención utilizando las Ecs. (2) - (4).

$$h = h(h) + \quad (2)$$

$$= \frac{((h\ h))}{((h\quad h))} \quad (3)$$

$$= h \quad (4)$$

Nuevamente pasamos la salida de la segunda capa BiLSTM (h) a un total capa conectada que produce la representación oculta (h) de una palabra. Pasamos esta representación oculta a una función softmax para medir la importancia de la palabra y obtener un peso de importancia normalizado ( ). Finalmente, la suma ponderada ( ) de la palabra representación produce la representación de características textuales atendidas. Aquí, , , y h son vectores inicializados aleatoriamente que aprenden conjuntamente durante el entrenamiento proceso.

Para hacer la dimensión final de la longitud 32, la salida de la capa de atención se pasa a través de una capa completamente conectada como se muestra en la Ec. (5) y produce la representación asistida del texto.

$$= (( ) +) \quad (5)$$

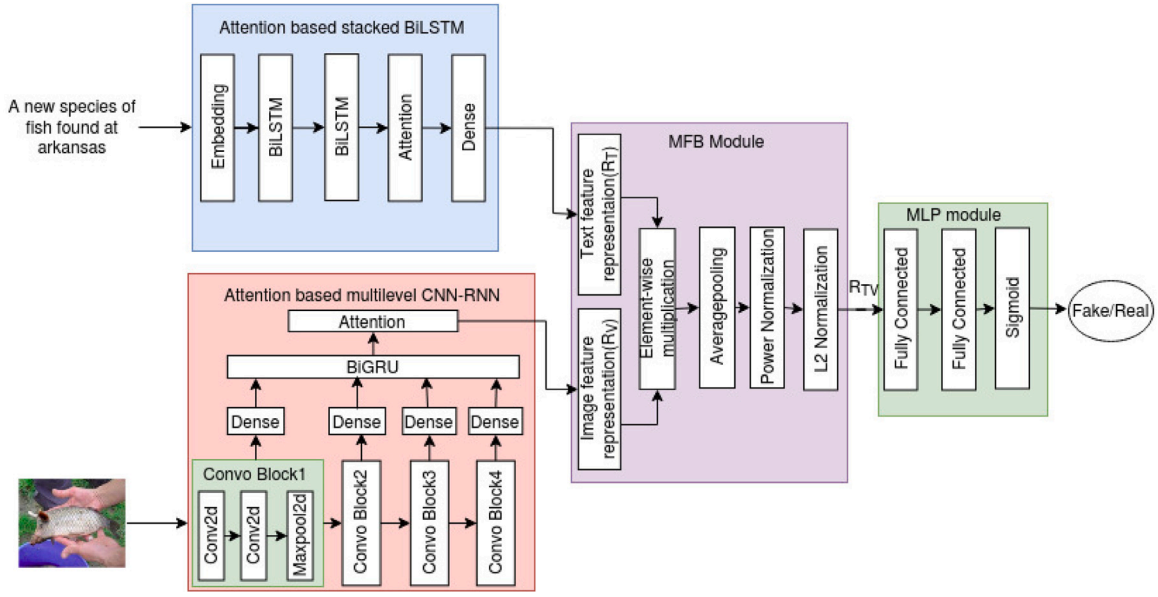

Fig. 3. AMFB: Agrupación bilineal factorizada multimodal basada en la atención para la detección multimodal de noticias falsas. En este modelo**Bi-LSTM apilado basado en la atención** extrae la representación textual de la característica, **CNN – RNN multinivel basado en la atención** extrae la representación de características visuales, **Agrupación bilineal factorizada multimodal** combina tanto textual como visual características para dar una representación compartida y **Perceptrón multicapa** finalmente clasifica la publicación de noticias como falsa o real.

4.2. CNN – RNN multinivel basado en la atención (ABM-CNN – RNN)

En general, las personas se sienten más tentadas hacia el contenido visual, ya que es más rápido y más fácil de capturar que el contenido textual. La mayoría de las publicaciones de noticias en la actualidad contienen algún contenido visual, como imágenes o videos. Por lo tanto, un modelo de extracción de características visuales preciso debería desempeñar un papel importante en la implementación de un detector multimodal de noticias falsas. Creamos CNN-RNN multinivel basado en la atención para la extracción de características visuales. Generalmente, CNN aprende características de alto nivel o características semánticas (características utilizadas por humanos para describir imágenes) a través de la abstracción capa por capa. Inicialmente, retiene elementos de bajo nivel como forma, línea, color, etc. Posteriormente, aprende características de alto nivel al enfocarse en objetos, acciones, etc., de la imagen. Las características semánticas de una imagen dependen en gran medida de las características de bajo nivel. La capa intermedia en CNN brinda información complementaria para la capa superior. Las características de bajo nivel a menudo se eclipsan y, a veces, se pierden debido a la representación de características semánticas de alto nivel. Literatura existente sobre clasificación de emociones de imágenes (Yang, Sun, Liang, Yang y Cheng, 2018) y detección de objetos salientes (Zhang, Wang, Lu, Wang y Ruan, 2017) han demostrado que la integración de características de nivel bajo y alto (semántico) proporciona mejores representaciones de características visuales que solo usando características de alto nivel.

Las imágenes de noticias falsas también muestran provocación emocional y algunos impactos visuales. Motivados por esta idea, diseñamos una subred CNN-RNN multinivel basada en la atención que captura mejores representaciones de características visuales semánticas mediante la integración de características de imagen de bajo y alto nivel. Inicialmente, la imagen con un tamaño (224,224) se da como entrada a una red CNN. Consideramos las dos capas de convolución bidimensionales con una función de activación ReLu y una capa de agrupación máxima bidimensional con un tamaño de agrupación (2,2) como un bloque CNN. Sobre cada salida de bloque, se agrega una capa completamente conectada para reducir la longitud de la entidad a 32. Ahora consideramos la salida de estos bloques como una secuencia y la pasamos a una Unidad Recurrente con Puerta bidireccional (BiGRU) para encontrar dependencias internas y secuenciales entre las características en ambas direcciones,
es decir, de nivel alto a nivel bajo y de nivel bajo a nivel alto. Al igual que las características textuales, la totalidad de las características visuales extraídas tampoco pueden tener la misma importancia, por lo que diseñamos un mecanismo de atención sobre la secuencia obtenida de BiGRU. Esto luego se pasa a través de un capa completamente conectada para producir la representación visual final ( ). Matemáticamente, se puede formular usando la Ec. (6) a la ecuación. (10):

$$= \quad (\quad 2 (\quad 2 (\quad - 1))) \tag{6}$$

$$= \quad ((\qquad) + \quad) \tag{7}$$

$$\begin{aligned} ?\overset{\rightarrow}{?} &= \overrightarrow{GRU}(??^{-}) \\ ?\overset{\leftarrow}{?} &= \overleftarrow{GRU}(??^{-}) \\ &= [?\overset{\rightarrow}{?}?, ?\overset{\leftarrow}{?}?] \\ &\in (1..4) \end{aligned} \tag{8}$$

$$= \quad ( ) \tag{9}$$

$$= \quad ((\quad) + \quad) \tag{10}$$

Aquí, , , y son los parámetros que se pueden aprender. En suma aria nuestro análisis revela que **ABS-BiLSTM** extrae las características textuales implícitas de alto nivel, mientras **ABM-CNN – RNN** extrae las características visuales multinivel. Los extractores de características propuestos son mejores que los mecanismos de extracción de características existentes, y también somos los primeros en utilizar un extractor de características novedoso en lugar de algunos métodos previamente entrenados. Este análisis resuelve nuestro primer objetivo de investigación.

4.3. Combinación bilineal factorizada multimodal (MFB)

Implementamos este componente del modelo propuesto para apoyar nuestra segundo objetivo. Después de obtener el texto final ( ) y visual) representaciones de características de ambas modalidades, las fusionamos usando el módulo MFB. Aquí, preferimos MFB sobre la concatenación estándar por las siguientes razones: (a). Es un desafío determinar el límite de las características extraídas obtenidas de las diferentes modalidades en la concatenación estándar.

(B). Después de la concatenación, las entidades se apilan una tras otra y por lo tanto, es posible que no descubra la correlación entre las representaciones de características de imagen y texto.

Estos dos problemas se pueden resolver de manera eficiente utilizando el módulo MFB. Este mecanismo de fusión maximiza la correlación entre las representaciones de características textuales y visuales. Supongamos que el rasgo textual el vector se denota como ( ) ∈ para texto y vector de características visuales como ( ) ∈ para una imagen. El modelo bilineal multimodal básico es entonces definido de acuerdo con la siguiente Eq. (11).

$$= \tag{11}$$

Dónde ∈ ∗ es una matriz de proyección. es la salida del modelo bilineal. La agrupación bilineal captura eficazmente la interacciones entre las dimensiones de la característica y simultáneamente introduce una gran cantidad de parámetros que acompañan a un alto nivel de cálculo costo y riesgo de sobreajuste. Para reducir el número de parámetros, en la ecuación. (11) se factoriza como dos matrices de rango bajo:

$$= \quad = \sum_{=1} \tag{12}$$

$$= 1 (\quad \circ \quad) \tag{13}$$

dónde es la dimensionalidad latente de las matrices factorizadas = [1,...,] ∈ ∗ y = [1,...,] ∈ ∗, ◦ es la multiplicación por elementos de dos vectores, 1 ∈ es un vector todo-uno. Para obtener la función de salida por Eq. (13), necesitamos aprender dos tensores de tres órdenes, = [1,....] ∈ ∗∗ y = [1,....] ∈ ∗∗ como pesos para la dimensión de salida. Puede reformularse aún más como bidimensional. matrices, ′ ∈ ∗ y ′ ∈ ∗ y se puede reescribir de la siguiente manera:

$$= \quad (\prime \circ \prime) \tag{14}$$

$$= \quad ( ) | \quad |0.5 \tag{15}$$

$$= \quad /|| \quad || \tag{dieciséis}$$

En resumen, podemos decir que nuestro mecanismo de fusión de características es diferente de los métodos de fusión existentes. La mayoría de los trabajos existentes se enfocaron en concatenar características textuales y visuales con el fin de obtener la representación compartida que muestra un desempeño muy limitado para la detección de noticias falsas. Siguiendo el segundo objetivo de investigación, diseñamos un mecanismo de fusión de características novedoso que maximiza la correlación entre características de texto e imagen y proporciona una alineación adecuada.

4.4. Perceptrón multicapa (MLP)

Diseñamos una subred de perceptrón multicapa con dos capas ocultas y una capa de salida con función de activación sigmoidea. Esta red de perceptrones multicapa toma características fusionadas como entrada. Los proyecta en el espacio objetivo de dos clases para producir la probabilidad de predicción final que decide si una publicación de noticias multimedia es falsa o real. En AMFB definimos la pérdida de entropía cruzada binaria entre las etiquetas original y predicha como función objetivo. Esto es matemáticamente formulado como se muestra en la Ec. (17) :

$$= - [\quad + (1 -) \quad (1 -)] \tag{17}$$

dónde *y* es la clase original y *pag* es la clase prevista de la publicación de noticias.

Todo lo anterior discutido cuatro componentes de la metodología propuesta en conjunto nos lleva a resolver nuestro tercer objetivo de investigación. En primer lugar, proporcionamos texto e imagen de una publicación de noticias como entrada, ABS-BiLSTM extrae características textuales, ABM-CNN-RNN extrae características visuales, MFB combina estas dos características y, finalmente, MLP determina si la noticia es falsa o real .

5. Conjunto de datos y experimentos

Para nuestros experimentos, utilizamos dos conjuntos de datos de referencia, *verbigracia.* Twitter y Weibo. Estos son los dos conjuntos de datos que los investigadores han utilizado para diseñar sistemas de detección de noticias falsas multimodales de alta calidad. Por lo tanto, para comparar con los trabajos anteriores, entrenamos nuestro modelo en estos dos conjuntos de datos de referencia del mundo real disponibles públicamente, es decir, Twitter.[1] y Weibo[2]

[1] https://github.com/MKLab-ITI/image-verification-corpus.

[2] https://drive.google.com/file/d/14VQ7EWPiFeGzxp3XC2DeEHi-BEisDINn/view?usp=sharing.

tabla 1

Distribución de datos de diferentes conjuntos de datos.

Conjunto de datos	Tren		Prueba		Imagen
	Falso	Verdadero	Falso	Verdadero	
Gorjeo	6841	5009	2564	1217	410
Weibo	3748	3783	1000	996	13274

Conjunto de datos de Twitter: El conjunto de datos de Twitter fue publicado por Boididou y col. (2015) como parte de la verificación del uso de multimedia en el desafío MediaEval. Este conjunto de datos consta de dos partes como un conjunto de entrenamiento y un conjunto de prueba. Los tweets de este conjunto de datos contienen texto, imágenes asociadas e información contextual. Los tweets en el conjunto de entrenamiento y prueba se han recopilado de los diferentes eventos y, por lo tanto, no hay superposición de eventos entre los conjuntos de entrenamiento y prueba. El conjunto de capacitación consta de 14483 tweets etiquetados con tres clases diferentes: falso, real y humorístico. Del total, 6848 tweets son falsos, 5001 tweets son reales y 2634 son de humor. El conjunto de prueba consta de 3781 tweets. Dado que el conjunto de prueba no contiene instancias de clases de humor, las ignoramos para nuestros experimentos. El conjunto de datos de Twitter contiene imágenes de conjuntos de entrenamiento de 360 y 50 imágenes de conjuntos de prueba. Usamos el 20% del conjunto de entrenamiento como conjunto de validación.

Conjunto de datos de Weibo: Weibo es un conjunto de datos chino multimodal. Weibo es un sitio web de microblogs en China que alienta a los usuarios a informar tweets sospechosos en Weibo. Esto luego es verificado como falso o real por el comité de usuarios reputados. Se recopilaron noticias falsas del conjunto de datos de Weibo desde mayo de 2012 hasta junio de 2016. Todos los tweets falsos, rastreados durante este tiempo y luego verificados por la Agencia de Noticias Xinhua, se consideran reales. La Agencia de Noticias Xinhua es una agencia de noticias autorizada en China. De acuerdo aJin y col. (2017), se han eliminado imágenes muy pequeñas y duplicadas para mantener la calidad del conjunto de datos. Las publicaciones sin imágenes también se han eliminado para que sean de naturaleza completamente multimodal. El conjunto de datos de Weibo consta de un total de 9527 publicaciones de noticias. Usamos 7531 como datos de entrenamiento y 1996 como datos de prueba. Los datos de entrenamiento se dividen en 90:10 para el entrenamiento y la validación del modelo. Este conjunto de datos también consta de 7954 imágenes falsas y 5320 reales.

tabla 1 muestra la distribución de datos completa para ambos conjuntos de datos.

Los conjuntos de datos discutidos anteriormente son los conjuntos de datos de referencia y están disponibles públicamente para la investigación de detección de noticias falsas. Inicialmente, hemos descargado estos conjuntos de datos de los respectivos repositorios. Lo hemos discutido en la sección de introducción, y solo consideramos el contenido de las noticias, es decir, texto e imagen, junto con las etiquetas individuales para realizar experimentos para nuestro modelo. Damos atributos de texto e imagen como entrada al modelo y mantenemos las etiquetas como verdad básica. En primer lugar, preprocesamos el texto y la imagen antes de ponerlos como entrada en la red. Para el texto, simbolizamos la oración, eliminamos la puntuación y las palabras vacías, y hacemos que todas las imágenes tengan el mismo tamaño. Convertimos el texto en formato vectorial usando la incrustación de texto rápido e imágenes en vectores usando imagen a matriz. Pasamos los vectores de texto al componente ABS-BiLSTM (discutido en la sección de metodología) del modelo propuesto para obtener la representación textual de las características. De manera similar, damos el vector de imagen en dos vectores como entrada en ABM-CNN-RNN (discutido en la sección de metodología) para obtener la representación de características visuales. Después de obtener estas dos representaciones, las pasamos al módulo MFB (descrito en la sección de metodología) del trabajo propuesto para obtener la representación de características multimodal compartida. Finalmente, damos esta representación compartida como entrada en el componente MLP (descrito en la sección de metodología) del modelo para proporcionar el resultado final como falso o real. Ahora, las etiquetas de verdad del terreno de los conjuntos de datos se comparan con las etiquetas predichas para calcular la pérdida. El conjunto de entrenamiento del conjunto de datos se usa para entrenar el modelo, y el conjunto de prueba se utiliza para validar el rendimiento del modelo. La pérdida se optimiza durante el proceso de entrenamiento y el modelo presenta la mejor hipótesis.

Tabla 2
Hiperparámetros utilizados para entrenar el modelo propuesto.

Parámetros	Gorjeo	Weibo
Longitud del texto	33	95
Tamaño de la imagen	(224,224,3)	(224,224,3)
Tamaño del lote	32	32
Optimizador	Adán (lr = 0,00001)	Adán (lr = 0,00005)
Épocas	100	100
Regularizador	L2 (0,5)	L2 (0,5)
Abandonar	0,2	0,2
Tamaño del filtro	(3,3)	(3,3)
Zancadas	(1,1)	(1,1)

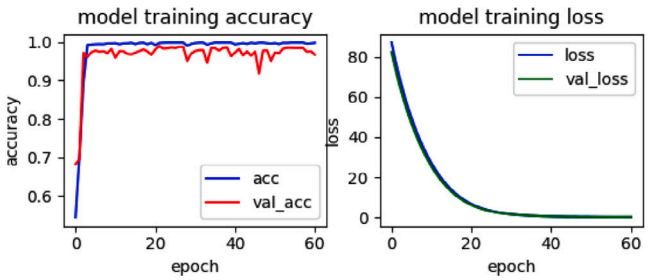

Figura 4. Curvas de aprendizaje para datos de Twitter.

5.1. Configuración experimental

En esta sección se analiza el mecanismo de inserción, el preprocesamiento de texto e imágenes, los diferentes hiperparámetros y los detalles de implementación. Todos los experimentos se llevan a cabo en un entorno de Python. Utilizamos las bibliotecas de Python Keras, NLTK, Numpy, Pandas y Sklearn para realizar los experimentos. Evaluamos el rendimiento del sistema en términos de exactitud, precisión, recuperación y puntuación F.

Eliminamos el identificador de Twitter, los signos de puntuación, los números, los caracteres especiales y las palabras cortas para la limpieza del texto de Twitter. Para los datos de Weibo, utilizamos el módulo Python de Jieba para la segmentación de datos. Para la extracción de características textuales, utilizamos la incrustación pre-entrenada de FastText de 300 dimensiones. Pasamos estos vectores de 300 dimensiones a la capa BiLSTM apilada, y luego la salida se entrega a una capa de atención que sigue a una capa densa. La primera y la segunda capa BiLSTM contienen 256 y 128 unidades, respectivamente, y la capa densa es 32.

Cambiamos el tamaño de todas las imágenes de tamaño (224,224,3), luego las pasamos a CNN – RNN multinivel basado en la atención. Aquí, usamos Conv2d con 64 unidades, (3,3) filtro de tamaño y (1,1) zancadas. Cada bloque CNN consta de dos capas CNN con activación ReLu y una capa Maxpool 2d con tamaño de grupo (2,2). La salida de cada bloque pasa a través de una capa completamente conectada de tamaño 32. La salida de los cuatro bloques se concatena y pasa a una capa GRU bidireccional de tamaño 32. Se presta atención a esta característica extraída y las características visuales atendidas son luego pasó a una capa completamente conectada de dimensión 32 con la función de activación ReLu para hacer que su tamaño sea igual a las características textuales. Usamos el módulo MFB para la fusión de características y lo pasamos a un perceptrón multicapa con dos capas ocultas de 32 y 16 con una función de activación ReLu y una capa de salida de tamaño uno con una función de activación Sigmoide. Aquí, usamos la entropía cruzada binaria como función de pérdida y optimizador de Adam. Entrenamos el modelo durante 100 épocas con 32 tamaños de lote y devoluciones de llamada de detención anticipada. Todos los hiperparámetros utilizados para entrenar el modelo propuesto se enumeran enTabla 2.

Implementamos algunos modelos de línea base basados en fuentes de información unimodales y multimodales para validar la efectividad de nuestro modelo propuesto.

Líneas de base unimodales. Definimos las siguientes líneas de base unimodales para las comparaciones: (a). Textual: para los modelos de línea de base textuales basados en características, utilizamos la incrustación de texto rápido previamente entrenado de 300 dimensiones para obtener el vector de características textuales y pasarlo a capas de tamaño LSTM bidireccionales 128. Estas representaciones de características textuales extraídas se entregan nuevamente a un MLP con dos capas ocultas con 64 y 32. La capa de salida de tamaño 2 con la función de activación Sigmoide detecta la publicación de noticias como falsa o real. (B). Visual: Para implementar el modelo visual, extraemos las características visuales con un modelo VGG19 previamente entrenado. Los vectores de características visuales extraídos son de tamaños 4096 que se alimentan nuevamente a una capa completamente conectada con un tamaño oculto 32 y una capa de salida con función de activación Sigmoide para la predicción.

Líneas de base multimodales. Para la multimodalidad, definimos las siguientes líneas de base:

(a). att-RNN: att-RNN (Jin y col., 2017) usa atención visual para detección multimodal de noticias falsas. El autor de este artículo originalmente utilizó la concatenación para la representación conjunta de texto e imagen. Usamos la multiplicación por elementos para obtener la representación conjunta de características multimodales. Para la incrustación, utilizamos un texto rápido de 300 dimensiones en lugar de la incrustación de word2vec. Los hiperparámetros son los mismos, como se menciona en el artículo original.

(B). EANN *: Event Adversarial Neural Network (EANN) (Wang et al., 2018) tiene tres componentes: el extractor de funciones, el discriminador de eventos y el detector de noticias falsas. En este modelo, el extractor de características extrae las características visuales y textuales invariantes del evento con la ayuda de un discriminador de eventos. Luego clasifica las publicaciones de noticias como falsas o reales usando estas funciones. Proponemos una variación de este modelo para la comparación de rendimiento, que tiene solo dos componentes: extractor de funciones y detector de noticias falsas. Todos los parámetros utilizados para entrenar este modelo son los mismos que los del modelo original.

(C). MVAE *: codificador automático variable multimodal (Khattar y col., 2019) entrena tres subredes para la detección de noticias falsas. Aquí, se entrenó un autocodificador variacional para obtener una mejor representación de características conjuntas textuales y visuales. La representación latente compartida se utilizó además para la clasificación. Aquí, construimos el modelo con codificador y partes del detector de noticias falsas para hacer una comparación justa. Entrenamos el modelo con los mismos hiperparámetros utilizados para entrenar el modelo original.

5.2. Resultados y análisis

El análisis comparativo de los modelos existentes y el modelo propuesto en dos conjuntos de datos diferentes se muestran en Tabla 3. Informamos la exactitud, precisión, recuperación y puntuación F1 de AMFB para clases reales y falsas. Los resultados muestran que nuestro enfoque propuesto produce un mejor rendimiento que los modelos de línea de base y de vanguardia existentes. Es evidente que el modelo visual ofrece un mejor rendimiento en comparación con el modelo textual. Esto puede deberse a que los textos a veces pueden contener información ruidosa y desestructurada, pero la imagen muestra una mejor evidencia. Se puede concluir a partir de los resultados que la incorporación de una imagen al texto es beneficiosa, ya que logra un mayor rendimiento en comparación con solo una imagen o un texto. Por lo tanto, este análisis demuestra haber logrado nuestro objetivo final de investigación.

Las curvas de aprendizaje para entrenamiento y validación muestran la pérdida y precisión del modelo propuesto (AMFB) para cada época. Higos. 4y 5 representan las curvas de aprendizaje del modelo propuesto para los conjuntos de datos de Twitter y Weibo, respectivamente. En ambas curvas de aprendizaje, la pérdida disminuye continuamente hasta una posición de equilibrio que muestra que el modelo aprende adecuadamente. Se considera que la curva de precisión de validación del conjunto de datos de Weibo fluctúa más que los datos de Twitter porque el tamaño de los datos de validación de Weibo es menor en comparación con los datos de validación de Twitter.

Para analizar cómo el modelo propuesto (AMFB) discrimina las noticias falsas de las noticias reales, empleamos la reducción de dimensionalidad utilizando Tdistributed Stochastic Neighbor Embedding (t-SNE) (Zhong, Li, Ma, Jiang y Zhao, 2017). Higos. 6y 7 muestran la proyección de las representaciones de características aprendidas por el modelo propuesto en el plano bidimensional para el conjunto de datos de Twitter y Weibo. Observamos que nuestro modelo obtiene una buena separabilidad para ambos conjuntos de datos. Superposición de instancias para

Tabla 3
Resultados de clasificación del modelo existente y propuesto en los conjuntos de datos de Twitter y Weibo.

Conjunto de datos	Modelo	Precisión	Noticias falsas			Noticias reales		
			Precisión	Recordar	Puntuación F	Precisión	Recordar	Puntuación F
Gorjeo	Textual	0.538	0,43	0,71	0,53	0,72	0,43	0,54
	Visual	0,645	0,52	0,59	0,55	0,74	0,68	0,71
	VQA	0,631	0,765	0,509	0,611	0.550	**0,794**	0,650
	Charla Neural	0,610	0,728	0,504	0.595	0.534	0,752	0,625
	att-RNN	0,664	0,749	0,615	0,676	0.589	0,728	0,651
	EANN *	0,741	0,69	0,55	0,61	0,76	0,85	0,81
	EANN	0,715	N / A	N / A	N / A	N / A	N / A	N / A
	MVAE *	0,724	0,62	0,64	0,63	0,79	0,77	0,78
	MVAE	0,745	0,801	0,719	0,758	0,689	0,777	0,730
	Spotfake	0,777	0,751	0.900	0,820	0,832	0,606	0,701
	AMFB	**0,883**	**0,89**	**0,95**	**0,92**	**0,87**	0,76	**0,81**
Weibo	Textual	0.593	0,62	0,50	0,55	0,58	0,69	0,63
	Visual	0,608	0,620	0,604	0,607	0,607	0,611	0,609
	VQA	0,736	0,797	0,634	0,706	0,695	0,838	0,760
	Charla Neural	0,726	0,794	0,713	0,692	0,684	0,840	0,754
	att-RNN	0,772	0,797	0,713	0,692	0,684	0,840	0,754
	EANN *	0,791	0,84	0,72	0,78	0,76	0,86	0,80
	EANN	0,827	N / A	N / A	N / A	N / A	N / A	N / A
	MVAE *	0,70	0,67	0,80	0,73	0,75	0,60	0,67
	MVAE	0,824	0,854	0,769	0,809	0,802	**0,875**	0,837
	Spotfake	**0.8923**	**0.902**	**0,964**	**0,932**	0,847	0,656	0,739
	AMFB	0,832	0,82	0,86	0,84	**0,85**	0,81	**0,83**

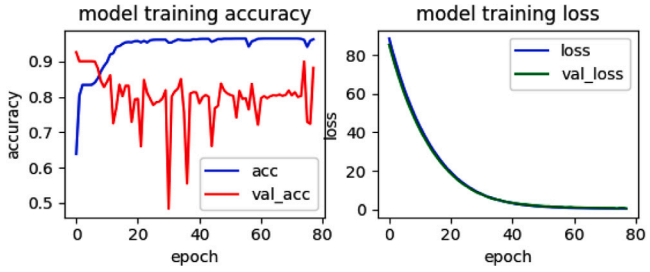

Figura 5. Curvas de aprendizaje para datos de Weibo.

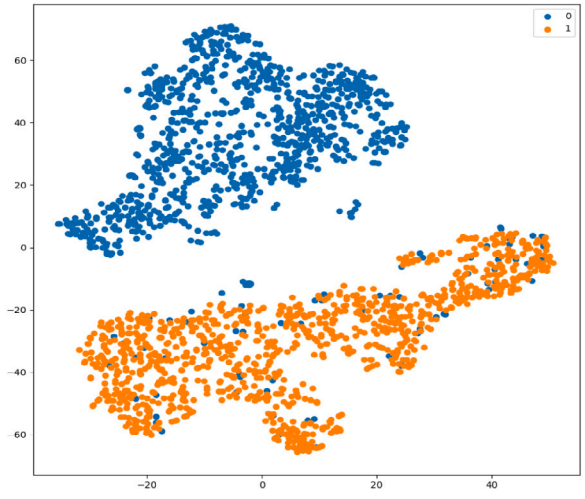

Figura 6. Proyección de representaciones de funciones para datos de Twitter.

el conjunto de datos de Weibo es más que el conjunto de datos de Twitter por dos razones: (i). La mayoría de las imágenes están más involucradas en el conjunto de datos de Weibo. (ii). Weibo es un conjunto de datos chino y, después de la segmentación, la longitud de algunas oraciones se vuelve mayor que la longitud de la oración del conjunto de datos de Twitter. El modelo no extrae mejores características textuales semánticas para oraciones complejas o muy largas. Estos inconvenientes se han explicado mejor en la sección de limitaciones.

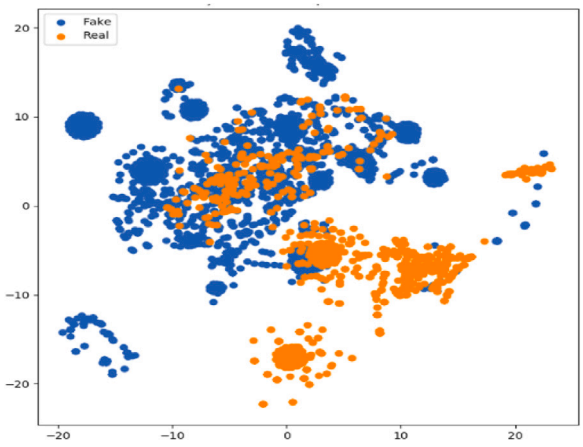

Figura 7. Proyección de representación de funciones para datos de Weibo.

Analisis cualitativo: Analizamos de cerca los resultados de los clasificadores para obtener mejores conocimientos. Figura 8 muestra dos tweets diferentes tomados del conjunto de datos de Twitter. El contenido del texto del primer tweet parece real y está clasificado como real por el modelo textual, pero tiene alteraciones en la imagen. En el segundo tweet, tanto el texto como la imagen son reales, y se clasifican como reales tanto por modelos textuales como visuales. Pero la imagen del segundo tuit no se tomó durante el terremoto de Nepal como se describe en el texto. Todos los tweets se muestran enFigura 8 son falsos y nuestro modelo propuesto (AMFB) predice que son falsos. Esto indica que el modelo propuesto extrae de manera efectiva características tanto visuales como textuales y decide si es falso o real.Figura 9 muestra algunos tweets reales tomados del conjunto de datos de Twitter. Estos tweets están clasificados correctamente por el modelo propuesto, mientras que los modelos de línea de base como los modelos textuales, visuales y existentes como att-RNN y EANN producen una predicción incorrecta. Por lo tanto, la observación anterior muestra que nuestro modelo propuesto supera el rendimiento textual, visual, att-RNN y EANN. El sistema de última generación, Spotfake, tampoco clasifica el primer tweet dirigido en

Figura 8, es decir, el tweet engañoso.

Comparación con el estado de las artes: En esta sección se describe el análisis comparativo del modelo propuesto y el estado actual del arte (Singhal y col., 2019). Los detalles de los resultados se muestran enTabla 3. En Spotfake (Singhal y col., 2019), BERT se ha utilizado para textuales

Figura 8. Tweets falsos correctamente clasificados por AMFB pero mal clasificados por modelo de modalidad única.

Figura 9. Algunos tuits reales correctamente clasificados por AMFB pero mal clasificados por el modelo existente.

extracción de características, que sin duda es un mecanismo poderoso. Usamos incrustación de texto rápido y LSTM apilado basado en la atención en lugar del codificador apilado de BERT para la extracción de características y, por lo tanto, nuestro modelo es menos complejo. Los datos de Twitter y Weibo son muy ruidosos y las oraciones no son sintáctica y semánticamente correctas. La mayoría de las palabras tampoco están completas en las oraciones, cediendo a los valores atípicos del vocabulario. BERT funciona mejor que fasttext; sin embargo, fasttext incluso puede manejarlo de una mejor manera. Por lo tanto, debido a la menor complejidad, utilizamos un mecanismo de texto rápido para incrustar. Spotfake utiliza una arquitectura VGG19 previamente entrenada para la extracción de características de imagen que extrae características de imagen generales y es posible que no extraiga las características de imagen invariantes. El estado de la técnica obtiene la representación de características conjuntas solo mediante la concatenación de características textuales y visuales que no dan una buena correlación entre imagen y texto. Entonces, las noticias engañosas (noticias en las que el contenido de texto del tweet describe la imagen capturada durante otro evento como se muestra en el segundo ejemplo deFigura 1) no se pueden clasificar correctamente. Para maximizar la correlación entre la representación de características textuales y visuales, nuestro modelo propuesto utiliza el módulo MFB, que brinda una mejor representación conjunta y ayuda al modelo a clasificar correctamente las publicaciones multimedia.

Nuestro modelo propuesto funciona mejor al tomar las medidas adecuadas para los problemas identificados por el modelo, Spotfake. Aunque el conjunto de datos de Twitter está desequilibrado y más de la mitad de las instancias pertenecen únicamente a un evento en particular, obtenemos una ganancia de precisión de aproximadamente un 9,8%. Spotfake informó una alta precisión para el conjunto de datos de Weibo. Incluso para un conjunto de datos equilibrado como Weibo, Spotfake informa una precisión, recuperación y puntuación F1 relativamente bajas para la clase real. Las imágenes en el conjunto de datos de Weibo son más complejas y el VGG19 previamente entrenado no puede capturar las características inherentes de la imagen de dominio específico de alto nivel que son la causa del resultado sesgado. Aunque nuestro modelo muestra una precisión inferior en comparación

Cuadro 4
Número de parámetros para AMFB y Spotfake.

Modelo	Numero de parámetros
AMFB	50,003,909
Spotfake	2,61,582,168

Para el modelo Spotfake, obtenemos valores de precisión, recuperación y puntuación F1 más equilibrados para ambas clases. Además, nuestro modelo no es tan complejo como Spotfake.

Para el análisis de complejidad de nuestro modelo propuesto en comparación con el Spotfake, analizamos el número de parámetros para cada modelo. Una gran cantidad de parámetros del modelo representan el modelo más complejo. Cuadro 4 muestra que el número de parámetros para el modelo propuesto (AMFB) es significativamente menor que el del estado de la técnica (Spotfake). Entonces, mediante la observación anterior, se muestra que el modelo propuesto funciona mejor incluso con menos complejidad.

Trascendencia Las noticias falsas, engañosas o falsas en las redes sociales pueden tener efectos sociales adversos significativos. La detección de noticias falsas en las redes sociales ha atraído recientemente la atención de investigadores y profesionales. Las investigaciones de este estudio tienen implicaciones tanto teóricas como prácticas. La mayoría de los trabajos anteriores exploraron fuentes unimodales (es decir, texto) para la detección de noticias falsas. Aunque existe una gran cantidad de investigaciones previas para detectar noticias falsas en las redes sociales, estos métodos aún no son completamente capaces de detectarlas en las primeras etapas o prevenirlas después de su difusión. Algunas de las técnicas existentes han intentado superar estas limitaciones utilizando solo el contenido de texto de las publicaciones de noticias. Dado que las publicaciones de noticias multimodales generan mucha atracción y atención que rápidamente pueden obligar a la población a creer, Las redes sociales están llenas de publicaciones de noticias multimodales hoy en día. Esto proporciona la motivación adecuada para detectar noticias falsas al aprovechar la información de múltiples modalidades.

Teóricamente, este estudio describe cómo extraer las características de diferentes modales y fusionar esas características para obtener la representación compartida que finalmente clasifica las noticias como falsas si el contenido de la imagen o el texto de una publicación de noticias contiene una información errónea. Hemos probado y validado empíricamente el papel de las imágenes en la detección de noticias falsas. En segundo lugar, estos hallazgos han descubierto nuevas facetas del intercambio y la detección de noticias falsas, proponiendo un marco integral basado en el aprendizaje profundo. Estas revelaciones contribuyen al conocimiento teórico en el dominio.

Prácticamente, podemos detectar las fake news en dos etapas: (i). inmediatamente después de introducir las fake news en la red, (ii). cuando las noticias falsas se propagan en la red. Nuestro objetivo principal es detectar las noticias falsas en los medios de la etapa temprana inmediatamente después de la introducción en la red y evitar que se propaguen solo aquí, pero también nuestro modelo monitorea la red regularmente para detectar noticias falsas. En cada etapa, solo toma como entrada el texto y la imagen, y en base a estos contenidos predice la veracidad de la noticia. Supongamos que la noticia se identifica como real en la etapa inicial, y en realidad es verdad, pero alguien ha realizado algún cambio en esa noticia en particular después. Dado que el modelo monitorea la red con regularidad, toma el texto y la imagen de esa noticia alterada como una nueva entrada y la detecta como falsa. Pero, si la noticia se identifica como real en la etapa inicial y en realidad es falsa, entonces el modelo no la detectará como falsa más adelante en la red hasta que alguien haya realizado cambios en el contenido de la noticia durante la propagación. Podemos concluir que si no hay alteraciones o cambios en el contenido de las noticias en ninguna etapa de la propagación, entonces no habrá ninguna diferencia en la decisión del modelo de detección de noticias falsas en las diferentes etapas. Nuestro modelo extrae las mejores características visuales y textuales sintácticas y semánticas y predice las noticias falsas con alta precisión, por lo que este es un caso raro. Podemos concluir que si no hay manipulación ni cambios en el contenido de las noticias en ninguna etapa de la propagación, entonces no habrá ninguna diferencia en la decisión del modelo de detección de noticias falsas en las diferentes etapas. Nuestro modelo extrae las mejores características visuales y textuales sintácticas y semánticas y predice las noticias falsas con alta precisión, por lo que este es un caso raro. Podemos concluir que si no hay alteraciones o cambios en el contenido de las noticias en ninguna etapa de la propagación, entonces no habrá ninguna diferencia en la decisión del modelo de detección de noticias falsas en las diferentes etapas. Nuestro modelo extrae las mejores características visuales y textuales sintácticas y semánticas y predice las noticias falsas con alta precisión, por lo que este es un caso raro.

Nuestros hallazgos muestran que la adopción de la estrategia de multimodalidad mejora el rendimiento de la detección de noticias falsas. Hemos demostrado la implementación práctica del marco de detección de noticias falsas multimodal basado en aprendizaje profundo en este estudio. No, toda la literatura existente se centra en la fusión de funciones multimodales para la detección de noticias falsas. Esto puede

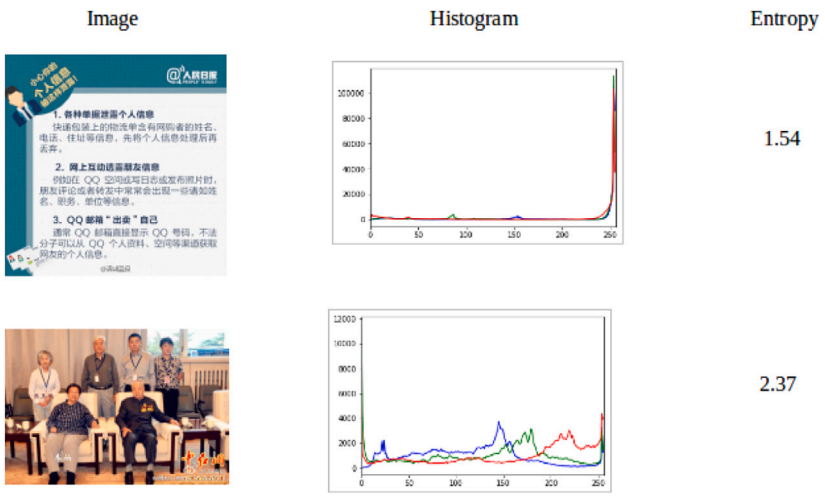

Figura 10. Análisis de complejidad de imágenes.

contribuyen en gran medida a mitigar la difusión y detección de noticias falsas, especialmente en las noticias con contenido de audio y video. Los resultados de la implementación muestran que nuestro modelo propuesto funciona mejor en comparación con el estado de la técnica. Los reguladores, investigadores, científicos, periodistas y organizaciones podrían beneficiarse de estos hallazgos y utilizarlos para prevenir y detectar la propagación de noticias falsas.

Limitaciones: Hemos intentado introducir y solucionar los problemas del estado del arte; sin embargo, nuestro modelo también tiene algunas limitaciones como *I)* El modelo no extrae características invariantes de imagen muy buenas de una imagen compleja. *ii)* Si la longitud del texto en la publicación de noticias es muy grande, a veces no se presta atención semántica a la última parte de la oración. *iii)* Nuestro modal propuesto proporciona una mejor correlación que el estado de la técnica, pero aún no captura la correlación semántica entre texto e imagen.

En Figura 10, analizamos la complejidad de algunas imágenes del Weibo conjunto de datos con entropía de Shannon (Khanzadi, Majidi y Akhtarkavan, 2017) y también trazar el histograma para mostrar la frecuencia de los valores de intensidad de píxeles. Nuestro modelo no funciona bien para las imágenes que tienen una entropía baja y una distribución no uniforme de la frecuencia de los valores de intensidad de los píxeles. Para el primer ejemplo enFigura 10, el modelo no extrae características invariantes útiles, mientras que el segundo modelo de ejemplo funciona bien.

Figura 11 muestra algunos ejemplos de conjuntos de datos de Twitter y Weibo, donde falla nuestro modelo propuesto. Ambos tweets muestran el texto atendido y su imagen correspondiente. En el primer ejemplo, la longitud del texto es muy grande, por lo que el modelo no detecta la atención en la última parte de la oración. Este tweet es falso, pero el modelo propuesto lo clasifica como real. El segundo ejemplo de tweet enFigura 11 explica mejor la tercera limitación. En este ejemplo, el modelo presta gran atención a*colapsa* y *terremoto* pero en la imagen no hay ningún objeto directo como *colapsa* y *terremoto*. En tales casos, nuestro modelo no puede encontrar las correlaciones semánticas y clasificarlas erróneamente como falsas.

Podemos abordar estas limitaciones para crear un gran impacto sostenible para resistir las noticias falsas en la red. También podemos incorporar las características del usuario, comportamiento del usuario, conocimiento de propagación, etc., que podrían impactar significativamente la detección de noticias falsas. Sin embargo, exige la creación de un conjunto de datos y un marco experimental adecuados.

6. Conclusión y labor futura

En este artículo, hemos propuesto un marco de aprendizaje profundo de extremo a extremo que toma la imagen y el texto de una publicación multimedia como entrada y detecta si esta publicación es falsa o real. Para extraer la característica textual, hemos diseñado un BiLSTM apilado basado en atención (ABS-BiLSTM). Para la extracción de características visuales, proponemos niveles múltiples basados en la atención

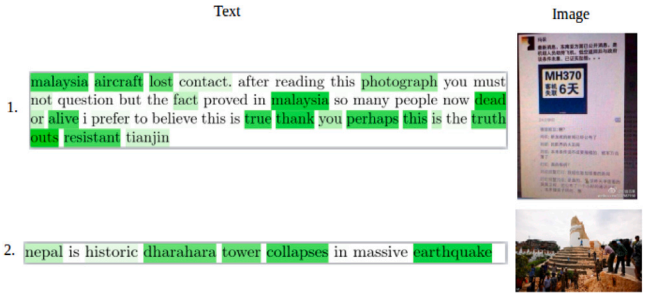

Figura 11. Algún ejemplo donde nuestro modelo falla.

CNN – RNN. La representación conjunta se obtiene a través del módulo MFB y se envía a un perceptrón multicapa (MLP) para detectar una publicación de noticias multimedia como falsa o real. Se han realizado extensos experimentos en dos conjuntos de datos multimodales disponibles públicamente. Los resultados experimentales muestran que el modelo propuesto funciona mejor para detectar noticias falsas en comparación con los modelos existentes.

Hay muchas posibilidades de investigar y ampliar este trabajo en el futuro. Aquí, enumeramos dos posibles direcciones de nuestra investigación futura: a) Las alineaciones semánticas entre texto e imagen pueden investigarse más a fondo para mejorar los mecanismos de fusión; (B). Hoy en día, muchas publicaciones de noticias contienen videos. Por tanto, una posible extensión sería la de incluir audios y videos.

Declaración de contribución de autoría de CRediT

Rina Kumari: Conceptualización, Metodología, Redacción de manuscrito. **Asif Ekbal:** Supervisión, Edición.

Declaración de intereses en competencia

Los autores declaran que no tienen intereses económicos en competencia o relaciones personales conocidas que pudieran haber influido en el trabajo informado en este documento.

Reconocimiento

Los autores agradecen el proyecto " HELIOS - Sistema de Observación y Elicitación del Odio, Hiperpartidista e Hiperpluralismo ", patrocinado por Wipro.

Referencias

Alkhodair, SA, Ding, SH, Fung, BC y Liu, J. (2020). Detectar noticias de última hora rumores de temas emergentes en las redes sociales. *Tratamiento y gestión de la información*, *57*(2), artículo 10 2018.

Antol, S., Agrawal, A., Lu, J., Mitchell, M., Batra, D., Lawrence Zitnick, C. y col. (2015). Vqa: Respuesta visual a preguntas. En*Actas de la conferencia internacional IEEE sobre visión artificial* (págs. 2425-2433).

Boididou, C., Andreadou, K., Papadopoulos, S., Dang-Nguyen, D.-T., Boato, G., Riegler, M. y col. (2015). Verificación del uso multimedia en medieval 2015 ..*Medieval*, *3*(3), 7.

Castillo, C., Mendoza, M. y Poblete, B. (2011). Credibilidad de la información en twitter. En *Actas de la 20a Conferencia Internacional sobre la World Wide Web* (págs. 675-684). ACM.

Chauhan, DS, Akhtar, MS, Ekbal, A. y Bhattacharyya, P. (2019). Consciente del contexto atención interactiva para análisis multimodal de sentimientos y emociones. En*Actas de la conferencia de 2019 sobre métodos empíricos en el procesamiento del lenguaje natural y la novena conferencia conjunta internacional sobre el procesamiento del lenguaje natural* (págs.5646-5656). Chauhan, H., Firdaus, M., Ekbal, A. y Bhattacharyya, P. (2019). Ordinal y atributo generación de respuesta consciente en un sistema de diálogo multimodal. En*Actas de la 57a reunión anual de la asociación de lingüística computacional* (págs. 5437-5447).

Iglesia, KW (2017). Word2vec.*Ingeniería del lenguaje natural*, *23*(1), 155-162. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. y Fei-Fei, L. (2009). Imagenet: una gran escala de la base de datos de imágenes jerárquicas. En*Conferencia IEEE de 2009 sobre visión por computadora y reconocimiento de patrones* (págs. 248-255). Ieee.

Devlin, J., Chang, M.-W., Lee, K. y Toutanova, K. (2018). Bert: Pre-entrenamiento de transformadores bidireccionales profundos para la comprensión del lenguaje. preimpresión arXivarXiv: 1810.04805.

Faustini, PHA y Covões, TF (2020). Detección de noticias falsas en múltiples plataformas e idiomas. *Sistemas Expertos con Aplicaciones*, Artículo 113503.

Fukui, A., Park, DH, Yang, D., Rohrbach, A., Darrell, T. y Rohrbach, M. (2016). Agrupación bilineal compacta multimodal para respuesta visual a preguntas y conexión a tierra visual. En*Conferencia sobre métodos empíricos en el procesamiento del lenguaje natural* (págs. 457-468). ACL.

Ghosal, D., Akhtar, MS, Chauhan, DS, Poria, S., Ekbal, A. y Bhattacharyya, P. (2018). Atención intermodal contextual para análisis de sentimiento multimodal. En *Actas de la conferencia de 2018 sobre métodos empíricos en el procesamiento del lenguaje natural* (págs. 3454–3466).

Gilda, S. (2017). Evaluación de algoritmos de aprendizaje automático para la detección de noticias falsas. En *2017 IEEE 15th conferencia de estudiantes sobre investigación y desarrollo (SCOReD)* (págs. 110-115). IEEE.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et Alabama. (2014). Redes generativas adversarias. En*Avances en los sistemas de procesamiento de información neuronal* (págs. 2672-2680).

Gravanis, G., Vakali, A., Diamantaras, K. y Karadais, P. (2019). Detrás de las señales: A estudio de evaluación comparativa para la detección de noticias falsas. *Sistemas Expertos con Aplicaciones*, *128* 201–213.

Huang, S., Cai, N., Pacheco, PP, Narrandes, S., Wang, Y. y Xu, W. (2018). Aplicaciones del aprendizaje de la máquina de vectores de apoyo (SVM) en la genómica del cáncer. *Genómica-proteómica del cáncer*, *15*(1), 41–51.

Huang, Y.-F. y Chen, P.-H. (2020). Detección de noticias falsas utilizando un aprendizaje conjunto modelo basado en algoritmos de búsqueda de armonía autoadaptativos. *Sistemas Expertos con Aplicaciones*, Artículo 113584.

Jin, Z., Cao, J., Guo, H., Zhang, Y. y Luo, J. (2017). Fusión multimodal con recurrente redes neuronales para la detección de rumores en microblogs. En*Actas de la 25a conferencia internacional ACM sobre multimedia* (págs. 795–816). ACM.

Jin, Z., Cao, J., Jiang, Y.-G. y Zhang, Y. (2014). Evaluación de la credibilidad de las noticias en microblog con un modelo de propagación jerárquico. En*Conferencia internacional IEEE 2014 sobre minería de datos* (págs. 230-239). IEEE.

Jin, Z., Cao, J., Zhang, Y., Zhou, J. y Tian, Q. (2016). Novedoso visual y estadístico características de imagen para la verificación de noticias de microblogs. *Transacciones IEEE en multimedia*, *19*(3), 598–608.

Karimi, H., Roy, P., Saba-Sadiya, S. y Tang, J. Noticias falsas de múltiples fuentes y clases múltiples detección. En*Actas de la 27a conferencia internacional de lingüística computacional* (págs. 1546-1557).

Karpathy, A. y Fei-Fei, L. (2015). Alineaciones visual-semánticas profundas para generar descripciones de imágenes. En*Actas de la conferencia IEEE sobre visión por computadora y reconocimiento de patrones* (págs. 3128-3137).

Khanzadi, P., Majidi, B. y Akhtarkavan, E. (2017). Una métrica novedosa para lo digital Evaluación de la calidad de la imagen mediante la complejidad de la imagen basada en la entropía. En*2017 IEEE 4th conferencia internacional sobre ingeniería e innovación basadas en el conocimiento (KBEI)* (págs. 0440–0445). IEEE.

Khattar, D., Goud, JS, Gupta, M. y Varma, V. (2019). MVAE: Variacional multimodal codificador automático para la detección de noticias falsas. En*La conferencia de la world wide web* (págs. 2915-2921). ACM.

Kim, J.-H., On, K.-W., Lim, W., Kim, J., Ha, J.-W. y Zhang, B.-T. (2016). Hadamard producto para agrupación bilineal de rango bajo. arXiv, arXiv – 1610.

Kwon, S., Cha, M., Jung, K., Chen, W. y Wang, Y. (2013). Características destacadas del rumor propagación en las redes sociales online. En*2013 IEEE 13th conferencia internacional sobre minería de datos* (págs. 1103-1108). IEEE.

Liu, Y., Jin, X. y Shen, H. (2019). Hacia la identificación temprana de rumores en línea basados en redes de memoria a corto plazo. *Tratamiento y gestión de la información*, *56*(4), 1457-1467.

Liu, Y., Pang, B. y Wang, X. (2019). Detección de spam de opinión mediante la incorporación de múltiples representación incrustada modal en un gráfico de revisión probabilística. *Neurocomputación*, *366*, 276–283.

Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, BJ, Wong, K.-F., et al. (2016). Detector rumores de microblogs con redes neuronales recurrentes. En *Ijcai* (págs. 3818–3824).

Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J. y Stein, B. (2018). Un estilometrico investigación sobre noticias falsas y hiperpartidistas. En*LCA (1)*.

Rashkin, H., Choi, E., Jang, JY, Volkova, S. y Choi, Y. (2017). Verdad de variar sombras: análisis del lenguaje en noticias falsas y verificación de hechos políticos. En*Actas de la conferencia de 2017 sobre métodos empíricos en el procesamiento del lenguaje natural* (págs. 2931-2937).

Reiter, E. y Dale, R. (1997). Construcción de sistemas aplicados de generación de lenguaje natural. *Ingeniería del lenguaje natural*, *3*(1), 57–87.

Rish, I. y col. (2001). Un estudio empírico del clasificador ingenuo de Bayes. En*IJCAI 2001 taller sobre métodos empíricos en inteligencia artificial* (págs. 41-46).

Rubin, V., Conroy, N., Chen, Y. y Cornwell, S. (2016). ¿Noticias falsas o verdad? utilizando señales satíricas para detectar noticias potencialmente engañosas. En*Actas del segundo taller sobre enfoques computacionales para la detección de engaños* (págs. 7-17).

Ruchansky, N., Seo, S. y Liu, Y. (2017). Csi: un modelo profundo híbrido para noticias falsas detección. En*Actas de la ACM de 2017 sobre la conferencia sobre gestión de la información y el conocimiento* (págs. 797-806). ACM.

Shaha, M. y Pawar, M. (2018). Transferir el aprendizaje para la clasificación de imágenes. En*2018 segunda conferencia internacional sobre electrónica, comunicaciones y tecnología aeroespacial (ICECA)* (págs. 656–660). IEEE.

Shen, H., Ma, F., Zhang, X., Zong, L., Liu, X. y Liang, W. (2017). Descubriendo lo social spammers desde múltiples puntos de vista. *Neurocomputación*, *225*, 49–57.

Shu, K., Cui, L., Wang, S., Lee, D. y Liu, H. (2019). Defender: noticias falsas explicables detección. En*25a conferencia internacional ACM SIGKDD sobre descubrimiento de conocimiento y minería de datos, KDD 2019* (págs. 395-405). Asociación para Maquinaria de Computación. Shu, K., Wang, S. y Liu, H. (2019). Más allá de los contenidos de las noticias: el papel del contexto social para la detección de noticias falsas. En *12a conferencia internacional de ACM sobre búsqueda web y minería de datos, WSDM 2019* (págs. 312-320). Association for Computing Machinery, Inc. Singhal, S., Shah, RR, Chakraborty, T., Kumaraguru, P. y Satoh, S. (2019). Spotfake: Un marco multimodal para la detección de noticias falsas. En*IEEE 2019 quinta conferencia internacional sobre big data multimedia (BigMM)* (págs. 39-47). IEEE.

Song, Y.-Y. y Ying, L. (2015). Métodos de árbol de decisión: aplicaciones para la clasificación y predicción. *Archivos de Psiquiatría de Shanghai*, *27*(2), 130.

ping Tian, D., et al. (2013). Una revisión sobre la extracción y representación de características de la imagen técnicas. *Revista internacional de ingeniería multimedia y ubicua*, *8*(4), 385–396.

Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., et al. (2018). Eann: Evento Redes neuronales adversas para la detección multimodal de noticias falsas. En*Actas de la 24a conferencia internacional acm sigkdd sobre descubrimiento de conocimiento y minería de datos* (págs. 849–857). ACM.

Wu, L., Rao, Y., Nazir, A. y Jin, H. (2020). Descubriendo características diferenciales: Adversario aprendizaje para la evaluación de la credibilidad de la información. *Ciencias de la información*, *516*, 453–473.

Yang, J., Sun, Y., Liang, J., Yang, Y.-L. y Cheng, M.-M. (2018). Entendiendo la imagen impresionante inspirado por las señales instantáneas de percepción humana. En*Trigésima segunda conferencia AAAI sobre inteligencia artificial*.

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A. y Hovy, E. (2016). Atención jerárquica redes para clasificación de documentos. En*Actas de la conferencia de 2016 del capítulo norteamericano de la asociación de lingüística computacional: tecnologías del lenguaje humano* (págs. 1480-1489).

Yu, Z., Yu, J., Fan, J. y Tao, D. (2017). Agrupación bilineal factorizada multimodal con aprendizaje de co-atención para la respuesta visual a preguntas. En*Actas de la conferencia internacional IEEE sobre visión artificial* (págs. 1821-1830).

Zhang, Z. (2016). Introducción al aprendizaje automático: k vecinos más cercanos.*Anales de Medicina traslacional*, *4*(11).

Zhang, H., Fang, Q., Qian, S. y Xu, C. (2019). Evento de conocimiento multimodal red de memoria para la detección de rumores en las redes sociales. *Actas de la 27a conferencia internacional ACM sobre multimedia* (págs. 1942-1951).

Zhang, P., Wang, D., Lu, H., Wang, H. y Ruan, X. (2017). Amuleto: agregando múltiples Funciones convolucionales de nivel para la detección de objetos destacados. En*Actas de la conferencia internacional IEEE sobre visión artificial* (págs.202-211).

Zhong, Z., Li, J., Ma, L., Jiang, H. y Zhao, H. (2017). Redes residuales profundas para clasificación de imágenes hiperespectrales. En*Simposio internacional de geociencias y teledetección del IEEE 2017 (IGARSS)* (págs. 1824-1827). IEEE.

Zubiaga, A., Kochkina, E., Liakata, M., Procter, R., Lukasik, M., Bontcheva, K., et Alabama. (2018). Clasificación de posturas de rumores conscientes del discurso en las redes sociales utilizando clasificadores secuenciales. *Tratamiento y gestión de la información*, *54*(2), 273–290.