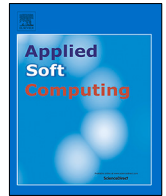




Listas de contenidos disponibles en ScienceDirect

Computación blanda aplicada

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

C-BiLSTM basado en la atención para la detección de noticias falsas

Tina Esther Trueman ^{a,*}, Ashok Kumar J. ^a, Narayanasamy P. ^B, Vidya J. ^a^a Departamento de Ciencia y Tecnología de la Información, Universidad de Anna, Chennai 600025, India^B Departamento de Ingeniería Eléctrica y Electrónica, Facultad de Tecnología de PSG, Coimbatore, India

información del artículo

Historia del artículo:

Recibido el 7 de octubre de 2020

Recibido en forma revisada el 31 de mayo de 2021

Aceptado el 4 de junio de 2021

On-line el 15 de junio de 2021

Palabras clave:

Detección de noticias falsas

Redes neuronales convolucionales Redes

neuronales recurrentes Memoria bidireccional

a corto y largo plazo Mecanismo de atención

abstracto

Las plataformas de redes sociales han transformado radicalmente la creación y difusión de noticias. Los usuarios pueden acceder fácilmente a estas noticias de una manera rápida y eficiente. Sin embargo, algunos usuarios pueden publicar contenido negativo y fraudulento en forma de comentarios o publicaciones. Dicho contenido puede constituir una amenaza para un individuo o una organización. Por lo tanto, la identificación de noticias falsas se ha convertido en un campo de investigación importante en el procesamiento del lenguaje natural (PNL). El principal desafío es determinar si la noticia es real o falsa. En este artículo, proponemos un enfoque de memoria convolucional bidireccional a largo plazo a corto plazo (AC-BiLSTM) basado en la atención para detectar noticias falsas y clasificarlas en seis categorías. La evaluación de nuestro enfoque propuesto en un conjunto de datos comparativos muestra una mejora significativa en la tasa de precisión en comparación con otros modelos de clasificación existentes. En particular, este trabajo contribuye al progreso en el campo de la detección de noticias falsas y confirma la viabilidad de nuestro enfoque propuesto para clasificar las noticias falsas en las redes sociales.

©2021 Elsevier BV Todos los derechos reservados.

1. Introducción

Los recientes desarrollos tecnológicos y la proliferación de Internet han tenido un impacto tremendo en las interacciones sociales. Las redes sociales se han convertido en una forma popular para que las personas interactúen y obtengan información. Las personas comparten sus actividades, intereses y opiniones personales a través de varias plataformas de redes sociales. Las redes sociales ofrecen un fácil acceso a la información con un costo menor y una información más rápida que las fuentes principales, como la transmisión y la impresión. Debido a estas características, muchas personas buscan noticias en las redes sociales en lugar de fuentes regulares de noticias como la televisión o el periódico. Como resultado, las noticias de las redes sociales están reemplazando rápidamente las fuentes de noticias clásicas. Aunque las redes sociales son muy útiles, el contenido de las noticias no siempre se puede ver como información confiable. La fuente de la noticia no está autenticada como lo estaría si se difundiera desde una fuente de noticias regular. Esto brinda la oportunidad de difundir noticias falsas. La información falsa se tergiversa como verdad. Los difusores de noticias falsas suelen tener un motivo oculto, como dañar la reputación de una persona o ganar dinero a través de afirmaciones falsas [1-3]. Las noticias falsas son peligrosas ya que pueden causar daños de diversas formas. Por ejemplo, puede difamar a una persona, organización o incluso provocar disturbios y protestas entre la gente. Puede afectar a sectores enteros de la sociedad. Por tanto, existe una necesidad creciente de clasificar las noticias como verdad o falsedad.

Hay bastantes sitios web que se dedican a verificar la veracidad de las afirmaciones hechas por las personas. PolitiFact. Esto se debe a que ciertas declaraciones se modifican sutilmente, lo que da una impresión falsa. Por lo tanto, un problema de clasificación de clases múltiples tiene sus propios desafíos. En el campo del aprendizaje automático, el descubrimiento de noticias falsas es uno de los problemas emergentes que han atraído la atención de investigadores de todo el mundo. Los algoritmos de aprendizaje automático necesitan una selección de características para clasificar los datos con precisión. Sin embargo, los algoritmos tradicionales de aprendizaje automático no tienen en cuenta la representación del contexto semántico. También, la identificación manual de la selección de características es una tarea de enormes proporciones. Las redes neuronales han logrado resultados razonables en comparación con otras metodologías de aprendizaje automático [4,5]. Recientemente, las noticias falsas se han detectado utilizando diversas técnicas de aprendizaje profundo como las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN). El RNN incluye LSTM (Redes de memoria a largo plazo a corto plazo), BiLSTM (Redes de memoria bidireccionales a largo y corto plazo), GRU (Unidades recurrentes cerradas (GRU) y BiGRU (Unidades recurrentes cerradas bidireccionales) [6,7].

Además, las técnicas de aprendizaje profundo tienen la ventaja de identificar las características por sí mismas. Estas técnicas identifican el significado de una palabra considerando su contexto. En particular, los mecanismos de atención se han convertido en una de las técnicas más poderosas en el procesamiento del lenguaje natural [8]. Se utilizan principalmente junto con redes neuronales recurrentes para predecir la información más relevante en la secuencia de entrada. Aunque la comunidad investigadora ha prestado considerable atención, una deficiencia

* Autor correspondiente.

Correos electrónicos: tina_trueman@auist.net (TE Trueman), jashokkumar83@auist.net (Ashok Kumar J.), dean.nc@psgtech.ac.in (Narayanasamy P.), vidyajeba@gmail.com (Vidya J.).

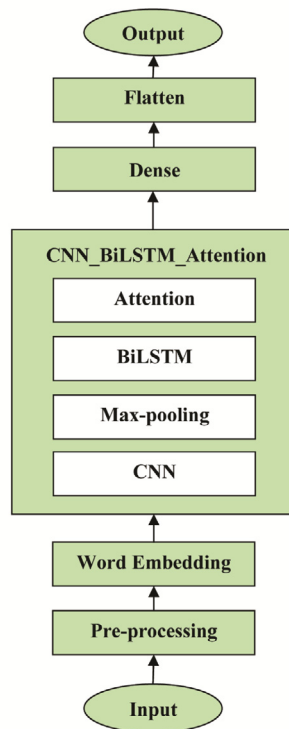


Figura 1. Modelo AC-BiLSTM para la detección de noticias falsas.

en la clasificación multiclase y basada en el contexto justifica una mayor investigación. Para abordar estos problemas, proponemos un modelo convolucional bidireccional de memoria a largo plazo a largo plazo basado en la atención para detectar noticias falsas. En particular, este trabajo ha sido motivado por la creencia de que la consideración del contexto mejora significativamente la precisión de la identificación de noticias falsas. Por tanto, este artículo contribuye a lo siguiente.

- Aborda el problema de detección de noticias falsas en un entorno de clases múltiples.
- Propone mecanismos de atención en conjunto con redes neuronales recurrentes bidireccionales convolucionales.
- Mejora la precisión de la detección de noticias falsas.

El resto de este documento está organizado de la siguiente manera. Sección2 analiza los trabajos relacionados en la detección de noticias falsas. La metodología de diseño del modelo propuesto se presenta en la Sección3. En la sección4, presentamos los resultados experimentales y las discusiones y su comparación significativa con los resultados de última generación (SOTA). Finalmente, concluimos este trabajo con posibles mejoras futuras en la Sección5.

2. Obras relacionadas

En esta sección, discutimos los métodos de aprendizaje profundo que se han utilizado en los trabajos existentes para la detección de noticias falsas. En particular, los métodos CNN y LSTM ya se han utilizado con éxito en nuevas tareas falsas, como lo ilustran los siguientes trabajos. Wang y col. [6] han presentado un método de detección de noticias falsas utilizando una CNN híbrida en un conjunto de datos comparativos. Específicamente, los autores combinaron los metadatos con texto. La CNN híbrida produjo más precisión que otros modelos existentes como SVM (Máquina de vectores de soporte), Regresión logística y BiLSTM. Sin embargo, el BiLSTM no funcionó bien debido al sobreajuste. Girgis y col. [7] presentó un trabajo para identificar noticias falsas utilizando modelos de aprendizaje profundo. Construyeron un modelo de clasificación híbrido

utilizando GRU y CNN. Los autores utilizaron el conjunto de datos LIAR para mostrar la eficiencia de su modelo. Han obtenido una precisión del 21,7%, 21,66% y 21,5% con los modelos GRU, LSTM y vanilla, respectivamente. Los autores buscaron aumentar la precisión proponiendo un modelo híbrido entre los métodos GRU y CNN sobre el mismo conjunto de datos.

Olivieri y col. [9] propuso una metodología que emplea características genéricas de tareas, junto con características textuales para detectar noticias falsas y clasificarlas en seis clases. Específicamente, las funciones genéricas de tareas se crearon en los metadatos adjuntos a las respuestas del motor de búsqueda de Google y del crowdsourcing. Rashkin y col. [10] sugirió un método basado en LSTM para detectar noticias falsas en el contexto de la verificación de hechos políticos. Para comparar las características lingüísticas de un texto poco confiable, compararon el idioma original de las noticias con el sarcasmo, el fraude y la propaganda. Khan et al. [11]. Demostraron que los modelos basados en redes neuronales entrenados por un conjunto de datos con menos de 100 artículos de noticias pueden lograr los mismos resultados que Naïve Bayes con N-gramos. Finalmente, realizaron un análisis temático, que expone la dificultad de conocer las noticias fraudulentas en política, salud e investigación.

Roy y col. [12] desarrolló modelos CNN y BiLSTM con una precisión de 42,89% y 42,65%, respectivamente. Los autores asignaron la representación obtenida de CNN y BiLSTM a un modelo MLP (perceptrón multicapa) para obtener la precisión de clasificación final (44,87%). Long y col. [13] propuso un modelo LSTM basado en la atención para integrar perfiles de oradores para detectar noticias falsas. Específicamente, los autores incluyeron perfiles de oradores como historial crediticio, título del orador, afiliación partidista y ubicación en el modelo de atención y la representación de entrada. Su estudio indicó que el modelo propuesto mejora la precisión en un 14,5% para el conjunto de datos de noticias falsas comparado. De la literatura anterior, es obvio que los métodos de aprendizaje profundo funcionan comparativamente mejor que los algoritmos clásicos de aprendizaje automático. Además, los métodos de aprendizaje profundo identifican automáticamente el mejor conjunto de características para un problema específico. Sin embargo, el concepto de mecanismo de atención juega un papel importante para detectar la información más relevante en la oración. Por lo tanto,

3. C-BiLSTM basado en la atención para la detección de noticias falsas

En esta sección, presentamos un modelo de memoria convolucional bidireccional a largo plazo a corto plazo (AC-BiLSTM) basado en la atención para detectar noticias falsas en un entorno multiclase. La arquitectura de este modelo se ilustra en Figura 1. Representa los componentes principales de nuestro modelo, como datos de entrada, preprocesamiento, incrustaciones de palabras, redes neuronales convolucionales (CNN), redes de memoria bidireccionales a largo plazo a corto plazo (BiLSTM) y mecanismo de atención. Analizamos cada uno de estos componentes de la siguiente manera.

3.1. Los datos de entrada

Demostramos el modelo propuesto utilizando el conjunto de datos LIAR que está disponible públicamente para detectar noticias falsas [6,14]. Este conjunto de datos contiene 12,8 K declaraciones que se recopilan de PolitiFact.com. Actualmente, es el mayor conjunto de datos disponible para la detección de noticias falsas. tabla 1 muestra la estructura del conjunto de datos. Contiene tres grupos de datos separados para entrenamiento, validación y prueba. Cada uno de estos grupos incluye 10 269, 1284 y 1283 declaraciones, respectivamente. Consisten en seis categorías etiquetadas, a saber, verdadero, medio verdadero, mayormente verdadero, apenas cierto, falso y pantalones en llamas.

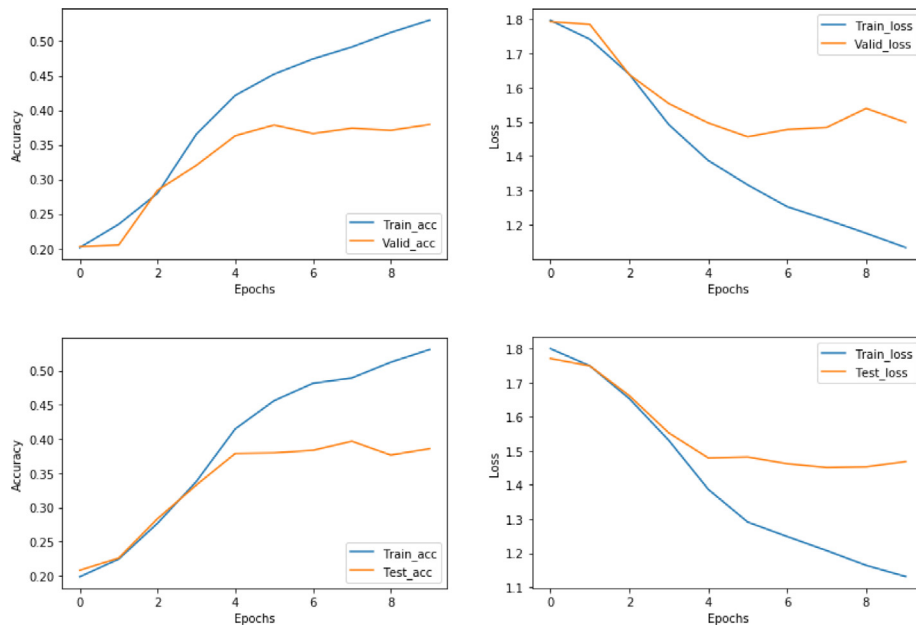


Figura 2. La curva de aprendizaje de AC-LSTM (F).

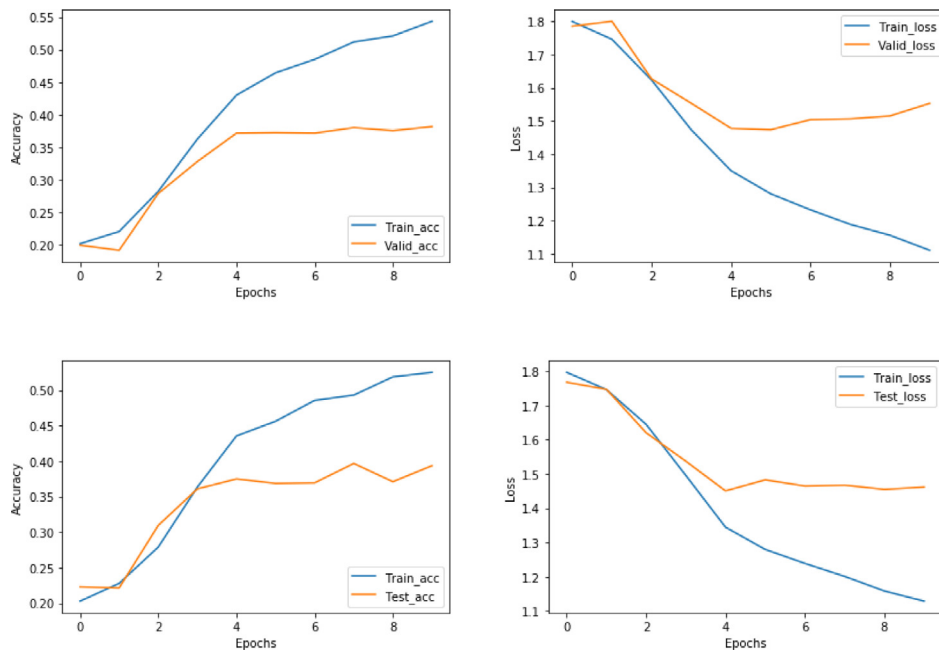


Fig. 3. La curva de aprendizaje de AC-LSTM (B).

tabla 1

Estructura del conjunto de datos de noticias falsas.

Categorías	Capacitación	Validación	Pruebas
Cierto	1683	169	211
La mitad de verdad	2123	248	267
Mayormente cierto	1966	251	249
Apenas cierto	1657	237	214
Falso	1998	263	250
Pantalones de fuego	842	116	92
Total	10,269	1284	1283

3.2. Preprocesamiento

Los datos de entrada se procesan previamente mediante varios pasos, como un conversor de casos, la tokenización y la eliminación de símbolos [7]. Primero, el conversor de mayúsculas y minúsculas se utiliza para convertir datos de texto en letras minúsculas. En segundo lugar, todos los símbolos se eliminan de los datos de texto. Finalmente, el proceso de tokenización se aplica para dividir el texto de entrada dado en secciones más pequeñas llamadas tokens. Estos pasos ayudan a transformar los datos de entrada dados en secuencias enteras.

3.3. Incrustaciones de palabras

Los datos obtenidos o preprocesados se pasan a la capa de incrustación de palabras. El objetivo principal de esta capa es generar

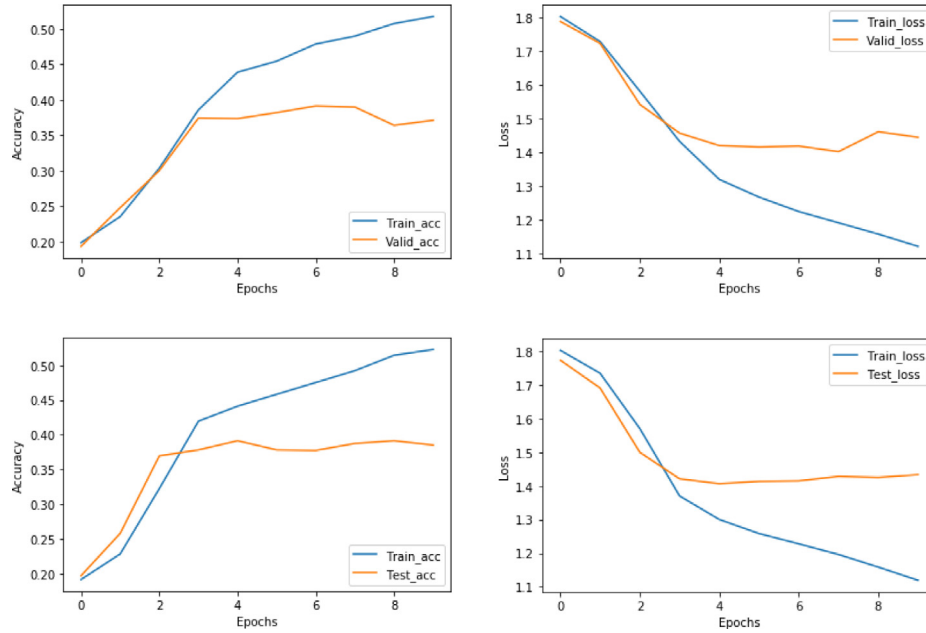


Figura 4. La curva de aprendizaje de AC-BiLSTM.

vectores de palabras para las tareas posteriores. Sin embargo, es una tarea desafiante construir vectores de palabras desde cero con miles de millones de palabras utilizando un gran conjunto de datos. Por lo tanto, utilizamos un modelo preentrenado que reduce la cantidad de tiempo necesario para limpiar, entrenar y procesar el modelo. Los modelos predefinidos (Word2Vec o GloVe) crean una palabra única que incorpora una representación de cada palabra en el vocabulario. En particular, utilizamos vectores de palabras GloVe con una dimensión fija. Se entrena como estadísticas de coincidencia palabra a palabra de un corpus de mil millones de palabras con un vocabulario de 400.000 palabras [7,15]. Específicamente, actualizamos los pesos aprendidos en nuestro modelo.

3.4. Redes neuronales convolucionales (CNN)

La CNN es un modelo de aprendizaje profundo que procesa información para realizar clasificación de imágenes, procesamiento automático del lenguaje natural, respuesta a preguntas, análisis de retroalimentación y tareas de resumen de texto con resultados notables [dieciséis-18]. Específicamente, CNN tiene una estructura única que facilita la detección de características de nivel superior. La capa convolucional es el principal bloque de trabajo de CNN y realiza la detección de características basada en coeficientes de matriz [19]. En particular, esta capa contiene un conjunto de núcleos o filtros. Estos filtros ayudan a procesar una parte limitada de la secuencia de entrada sobre todos los datos de entrada a través de una función de activación llamada ReLU (Unidad lineal rectificadora). El ReLU es la elipse de la unidad lineal corregida [20] se establece principalmente en cero en la red para eliminar valores negativos del mapa de activación. Por lo tanto, esta función de activación es computacionalmente más eficiente para resolver el problema del gradiente invisible que las funciones sigmoidea y tanh.

3.5. Capa de agrupación máxima

Las características detectadas a través de la capa convolucional se pasan a una capa de agrupación. La capa de agrupación resume las características detectadas en la entrada y produce un nuevo mapa de características. También se denomina método de muestreo descendente no lineal [21,22]. Este método reduce progresivamente la dimensión de la característica para reducir el número de parámetros en la red computacional. En este documento, usamos una capa de agrupación máxima para seleccionar el valor máximo en un mapa de características con la ayuda del tamaño del filtro.

3.6. Redes bidireccionales de memoria a largo plazo a corto plazo (BiLSTM)

Una red neuronal recurrente (RNN) se utiliza principalmente para procesar datos secuenciales. Almacena las partes relevantes de los datos de entrada y usa esta información para predecir la salida futura. El RNN contiene celdas de memoria, que se utilizan para almacenar la información más relevante de las entradas pasadas. Además, no puede manejar las dependencias a largo plazo. Por lo tanto, se introduce un tipo especial de RNN para resolver las dependencias a largo plazo llamado red de memoria a largo plazo a corto plazo (LSTM) [23,24]. Esto se implementa mediante tres conceptos de activación [dieciséis,25,26], es decir, entrada (YO_G), producción (OG), y olvídate de las puertas (FG). Estas puertas controlan el flujo de información (lectura, escritura y reinicio) sobre el gradiente junto con el estado oculto candidato (CHS), estado actual (CS) y secuencia oculta (HS) como en las Ecs. (1)-(6). Específicamente, la red LSTM procesa información unidireccionalmente, ya sea de izquierda a derecha o de derecha a izquierda. Conduce al problema de obtener información futura. En este caso, se introduce un LSTM bidireccional para aprender la secuencia de entrada original desde el principio hasta el final y desde el final hasta el final.

comenzando como en Eq. (7).

$$YO_G = \sigma(W_{YO_G}X_t + R_{YO_G}h_{t-1} + B_{YO_G}) \quad (1)$$

$$OG = \sigma(W_{OG}X_t + R_{OG}h_{t-1} + B_{OG}) \quad (2)$$

$$FG = \sigma(W_{FG}X_t + R_{FG}h_{t-1} + B_{FG}) \quad (3)$$

$$CHS_t = \tanh(W_{CHS}X_t + R_{CHS}h_{t-1} + B_{CHS}) \quad (4)$$

$$CS_t = FG_t \otimes CS_{t-1} + YO_G \otimes HS_t \quad (5)$$

$$HS_t = OG_t \otimes \tanh(CS_t) \quad (6)$$

$$y_t = \overrightarrow{HS}_t \overleftarrow{HS}_t \quad (7)$$

Dónde W_{YO_G} , W_{OG} , W_{FG} , y W_{CHS} se refieren a las matrices de peso de la entrada actual X_t . R_{YO_G} , R_{OG} , R_{FG} , y R_{CHS} se conocen como las matrices de peso del estado anterior h_{t-1} . B_{YO_G} , B_{OG} , B_{FG} , y B_{CHS} se denotan como el valor de sesgo. y_t representa la salida de las unidades LSTM hacia adelante y hacia atrás LSTM.

Tabla 2

Modelo de hiperparámetros.

Hyper parámetros	Tamaño	Hyper parámetros	Tamaño
Dimensión del guante	100	Unidades BiLSTM	150
Longitud de la secuencia de entrada	150	Activación de atención	Sigmoideo
Tamaño del vocabulario	10,000	Capa densa	32
Tasa de deserción escolar	0,5	Tamaño del lote	64
Filtros	300	Activación de salida	Softmax
Tamaño de ventana	5	Optimizador	Adán (0,001)
Tamaño de la agrupación	4	Función de pérdida	Entropía cruzada categórica

Tabla 3

Matriz de confusión de formación y validación para AC-LSTM (F).

Clase	Capacitación						Validación					
	0	1	2	3	4	5	0	1	2	3	4	5
0	1428	47	240	47	96	140	126	24	35	12	56	10
1	146	530	149	85	746	27	41	47	21	13	43	4
2	328	69	1051	73	86	50	54	24	93	20	40	6
3	243	127	250	771	720	12	30	34	26	73	73	4
4	sesenta y cinco	137	88	63	1608	5	50	25	29	21	122	4
5	163	2	55	2	2	618	41	11	11	3	24	26

Cuadro 4

Matriz de confusión de pruebas y entrenamiento para AC-LSTM (F).

Clase	Capacitación						Pruebas					
	0	1	2	3	4	5	0	1	2	3	4	5
0	1514	44	144	121	74	101	133	10	21	51	31	4
1	167	587	100	322	490	17	58	43	26	38	46	0
2	419	52	861	236	60	29	45	8	61	48	46	6
3	265	90	116	1285	357	10	58	10	18	117	60	4
4	107	123	47	298	1388	3	50	22	20	47	110	0
5	167	2	45	7	2	619	22	6	8	19	6	31

Cuadro 5

Matriz de confusión de formación y validación para AC-LSTM (B).

Clase	Capacitación						Validación					
	0	1	2	3	4	5	0	1	2	3	4	5
0	1384	104	176	106	163	sesenta y cinco	123	36	7	43	53	1
1	95	719	66	121	669	13	34	46	19	29	40	1
2	324	98	873	177	165	20	45	35	74	48	34	1
3	170	205	96	991	656	5	40	43	9	110	46	0
4	46	215	20	77	1605	3	44	28	15	40	123	1
5	193	4	57	11	4	573	40	20	3	18	21	14

Tabla 6

Matriz de confusión de pruebas y entrenamiento para AC-LSTM (B).

Clase	Capacitación						Pruebas					
	0	1	2	3	4	5	0	1	2	3	4	5
0	1690	58	39	116	32	63	132	31	10	52	23	2
1	218	662	56	286	442	19	54	55	18	43	40	1
2	611	79	655	236	44	32	44	24	50	57	31	8
3	346	155	44	1271	294	13	64	28	9	129	35	2
4	119	199	20	360	1265	3	49	32	9	53	106	0
5	226	7	25	12	1	571	20	6	1	26	6	33

Tabla 7

Matriz de confusión de formación y validación para AC-BiLSTM.

Clase	Capacitación						Validación					
	0	1	2	3	4	5	0	1	2	3	4	5
0	1439	109	113	75	89	173	144	64	6	12	30	7
1	112	1079	55	123	290	24	40	69	15	15	24	6
2	490	162	741	137	51	76	58	66	sesenta y cinco	20	18	10
3	320	524	98	881	278	22	56	79	8	71	27	7
4	101	799	26	111	925	4	57	67	6	14	102	5
5	178	5	29	11	0	619	45	33	1	5	7	25

Tabla 8

Matriz de confusión de pruebas y entrenamiento para AC-BiLSTM.

Clase	Capacitación						Pruebas					
	0	1	2	3	4	5	0	1	2	3	4	5
0	1533	32	219	114	39	61	177	2	12	37	20	2
1	192	443	102	434	498	14	69	38	20	46	38	0
2	452	25	978	146	31	25	69	9	74	39	19	4
3	269	34	265	1359	187	9	97	14	27	106	23	0
4	160	55	120	508	1119	4	87	12	23	56	71	0
5	242	3	62	13	0	522	31	4	12	14	3	28

Cuadro 9

El rendimiento del AC-LSTM (F).

Clase	Capacitación			Validación			Capacitación			Pruebas		
	PAG	R	F1	PAG	R	F1	PAG	R	F1	PAG	R	F1
0	0,60	0,71	0,65	0,36	0,48	0,41	0,57	0,76	0,65	0,36	0,53	0,43
1	0,58	0,31	0,41	0,28	0,28	0,28	0,65	0,35	0,45	0,43	0,20	0,28
2	0,57	0,63	0,60	0,43	0,39	0,41	0,66	0,52	0,58	0,40	0,29	0,33
3	0,74	0,36	0,49	0,51	0,29	0,37	0,57	0,61	0,59	0,37	0,44	0,40
4	0,49	0,82	0,62	0,34	0,49	0,40	0,59	0,71	0,64	0,37	0,44	0,40
5	0,73	0,73	0,73	0,48	0,22	0,31	0,79	0,74	0,76	0,69	0,34	0,45
MAC	0,62	0,60	0,60	0,40	0,36	0,36	0,64	0,61	0,61	0,44	0,37	0,38
MIC	0,58	0,58	0,58	0,38	0,38	0,38	0,61	0,61	0,61	0,39	0,39	0,39
Washington	0,61	0,58	0,58	0,40	0,38	0,38	0,62	0,61	0,60	0,41	0,39	0,38

Tabla 10

El rendimiento del AC-LSTM (B).

Clase	Capacitación			Validación			Capacitación			Pruebas		
	PAG	R	F1	PAG	R	F1	PAG	R	F1	PAG	R	F1
0	0,63	0,69	0,66	0,38	0,47	0,42	0,53	0,85	0,65	0,36	0,53	0,43
1	0,53	0,43	0,47	0,22	0,27	0,24	0,57	0,39	0,47	0,31	0,26	0,28
2	0,68	0,53	0,59	0,58	0,31	0,41	0,78	0,40	0,52	0,52	0,23	0,32
3	0,67	0,47	0,55	0,38	0,44	0,41	0,56	0,60	0,58	0,36	0,48	0,41
4	0,49	0,82	0,61	0,39	0,49	0,43	0,61	0,64	0,63	0,44	0,43	0,43
5	0,84	0,68	0,75	0,78	0,12	0,21	0,81	0,68	0,74	0,72	0,36	0,48
MAC	0,64	0,60	0,61	0,45	0,35	0,35	0,64	0,59	0,60	0,45	0,38	0,39
MIC	0,60	0,60	0,60	0,38	0,38	0,38	0,60	0,60	0,60	0,39	0,39	0,39
Washington	0,62	0,60	0,59	0,43	0,38	0,38	0,62	0,60	0,59	0,42	0,39	0,39

Cuadro 11

El rendimiento del AC-BiLSTM.

Clase	Capacitación			Validación			Capacitación			Pruebas		
	PAG	R	F1	PAG	R	F1	PAG	R	F1	PAG	R	F1
0	0,55	0,72	0,62	0,36	0,55	0,43	0,54	0,77	0,63	0,33	0,71	0,45
1	0,40	0,64	0,49	0,18	0,41	0,25	0,75	0,26	0,39	0,48	0,18	0,26
2	0,70	0,45	0,55	0,64	0,27	0,38	0,56	0,59	0,57	0,44	0,35	0,38
3	0,66	0,41	0,51	0,52	0,29	0,37	0,53	0,64	0,58	0,36	0,40	0,39
4	0,57	0,47	0,51	0,49	0,41	0,44	0,60	0,57	0,58	0,41	0,29	0,34
5	0,67	0,74	0,70	0,42	0,22	0,28	0,82	0,62	0,71	0,82	0,30	0,44
MAC	0,59	0,57	0,56	0,44	0,36	0,36	0,63	0,57	0,58	0,47	0,37	0,38
MIC	0,55	0,55	0,55	0,37	0,37	0,37	0,58	0,58	0,58	0,39	0,39	0,39
Washington	0,58	0,55	0,55	0,45	0,37	0,38	0,61	0,58	0,57	0,43	0,39	0,44

3.7. Mecanismo de atención

La capa de atención crea un vector de contexto para los vectores de entrada aprendidos. Desempeña un papel influyente en la traducción automática, el reconocimiento de imágenes, el resumen de texto, la clasificación de texto y los sistemas de respuesta a preguntas. En este documento, la capa de atención se construye en la parte superior de la red BiLSTM para actualizar los pesos. Específicamente, la capa de atención asigna selectivamente mayor peso a las palabras más relevantes e importantes de la secuencia de entrada. La ventaja de este mecanismo es preservar secuencias de entrada más largas [8,27].

3.8. Capa densa

La capa densa o completamente conectada realiza una operación lineal que conecta todos los valores de entrada a cada neurona en la siguiente capa.

Cuadro 12

La precisión media del modelo propuesto.

Modelos	Capacitación	Validación	Capacitación	Pruebas
AC-LSTM (F)	0.3963	0.3243	0.3923	0.3351
AC-LSTM (B)	0.4022	0.3250	0.3987	0.3389
AC-BiLSTM	0.4007	0.3384	0.4065	0.3513

por peso. La función de esta capa es proporcionar todas las combinaciones de características de la capa anterior a la siguiente [dieciséis,22,28].

3.9. Aplanar capa

La capa de aplanamiento transforma la dimensión de la capa densa en una sola dimensión sin afectar el tamaño del lote. Por ejemplo, las características dimensionales 2D se transforman en

Cuadro 13

Comparación de resultados del modelo propuesto.

Autores	Modelos	Precisión		Puntuación F1		Micro F1		Macro F1		F1 ponderado	
		Válido	Prueba	Válido	Prueba	Válido	Prueba	Válido	Prueba	Prueba válida	
Wang y col. [6]	Mayoria	0,204	0,208	-	-	-	-	-	-	-	-
	SVM	0,258	0,255	-	-	-	-	-	-	-	-
	Log_Reg	0,257	0,247	-	-	-	-	-	-	-	-
	BiLSTM	0,223	0,233	-	-	-	-	-	-	-	-
	CNN	0,260	0,270	-	-	-	-	-	-	-	-
	CNN híbrido	0,247	0,274	-	-	-	-	-	-	-	-
Olivieri y col. [9]	NN	-	-	0,273	-	-	-	-	-	-	-
	SVM	-	-	0,308	-	-	-	-	-	-	-
Rashkin y col. [10]	Mayoria BL	-	-	-	-	-	-	-	0,06	-	-
	NB (texto + LIWC)	-	-	-	-	-	-	0,21	0,17	-	-
	MaxEnt (texto + LIWC)	-	-	-	-	-	-	0,21	0,22	-	-
	LSTM (texto + LIWC)	-	-	-	-	-	-	0,22	0,19	-	-
	LSTM (texto)	-	-	-	-	-	-	0,21	0,20	-	-
Girgis y col. [7]	SVM	-	0,255	-	-	-	-	-	-	-	-
	Log_Reg	-	0,247	-	-	-	-	-	-	-	-
	BiLSTM	-	0,233	-	-	-	-	-	-	-	-
	CNN	-	0,270	-	-	-	-	-	-	-	-
	Vanila	-	0,215	-	-	-	-	-	-	-	-
	GRU	-	0,217	-	-	-	-	-	-	-	-
	LSTM	-	0,217	-	-	-	-	-	-	-	-
Roy y col. [12]	BiLSTM	-	0,427	-	0,41	-	-	-	-	-	-
	CNN	-	0,429	-	0,42	-	-	-	-	-	-
	RNN_CNN	-	0,449	-	0,43	-	-	-	-	-	-
Yunfei Long [13]	Base LSTM	0,250	0,255	-	-	-	-	-	-	-	-
	LSTM híbrido	0,407	0,415	-	-	-	-	-	-	-	-
Alhindi y col. [29]	Log_Reg	-	-	0,38	0,37	-	-	-	-	-	-
	SVM	-	-	0,35	0,35	-	-	-	-	-	-
	BiLSTM	-	-	0,34	0,32	-	-	-	-	-	-
	P_Bi_LSTM	-	-	0,37	0,36	-	-	-	-	-	-
Propuesto	AC-LSTM (F)	0,324	0,335	0,36	0,38	0,38	0,39	0,36	0,38	0,38	0,38
	AC-LSTM (B)	0,325	0,339	0,35	0,39	0,38	0,39	0,35	0,39	0,38	0,39
	AC-BiLSTM	0,338	0,351	0,36	0,38	0,37	0,39	0,36	0,38	0,38	0,37

una característica dimensional unidimensional. La salida de esta capa se aplica a la capa de salida softmax [dieciséis,28].

3.10. Capa Softmax

La capa softmax mide la probabilidad de todas las clases posibles en un problema de varias clases. Transforma los valores de entrada aprendidos entre cero y uno. Esta asignación de probabilidad ayuda al modelo a converger rápidamente [dieciséis,22,28]. La función softmax es definido como sigue en Eq. (8).

$$\text{softmax}(z)_i = \frac{e^{m_{zi}}}{\sum_{j=1}^k m_{zj}} \quad (8)$$

Dónde z se refiere a los elementos del vector de entrada que toma cualquier valor real y k se refiere al número de clases. El denominador de la función normaliza todos los valores de salida en el rango de 0 a 1. El valor de probabilidad más alto es la salida de la clase.

4. Resultados y discusión

En esta sección, presentamos el desempeño del modelo AC-BiLSTM propuesto usando las métricas de evaluación estándar como matriz de confusión, exactitud, precisión, recuperación, puntaje F1, promedio macro, promedio micro y promedio ponderado [30]. Una matriz de confusión o una matriz de errores es una tabla que representa el desempeño de un clasificador basado en verdaderos positivos, falsos positivos, falsos negativos y verdaderos negativos. La precisión mide el total de documentos correctamente identificados de todos los documentos. Precision calcula la capacidad de un clasificador para identificar todos los documentos relevantes. El recuerdo mide la capacidad de un clasificador de devolver solo documentos relevantes. La puntuación F1 calcula la media armónica de

precisión y recuerdo. La puntuación media macro calcula la métrica para cada clase con ponderaciones iguales. No considera el problema del desequilibrio de clases. El puntaje micropromedio mide la métrica para cada instancia con ponderaciones iguales basadas en el total de verdaderos positivos, falsos positivos y falsos negativos. La puntuación media ponderada calcula métricas para cada clase en función del número de instancias reales. Es lo mismo que el puntaje macro promedio, pero considera el problema del desequilibrio de clases. En este documento, utilizamos un conjunto de datos de referencia LIAR para la detección de noticias falsas. El conjunto de datos LIAR contiene 10,269 instancias para entrenamiento, 1284 instancias para validación y 1283 instancias para prueba. Luego empleamos técnicas de preprocesamiento, como eliminación de caracteres y símbolos, conversión de caracteres y tokenización para obtener la calidad de los documentos de entrada. Estos documentos se transforman en una secuencia de números enteros. Posteriormente, la incrustación de palabras previamente entrenadas de GloVe se utiliza para generar vectores de palabras para cada palabra en el documento de capacitación. Estos vectores de palabras se suministran como secuencias de entrada al modelo AC-BiLSTM propuesto. Tabla 2 muestra los hiperparámetros del modelo. El modelo se implementa en un entorno de CPU con un sistema de Windows utilizando el software Anaconda con varias bibliotecas de aprendizaje automático como Keras y scikit. El rendimiento del modelo propuesto se evalúa con las métricas mencionadas anteriormente en el pase hacia adelante, el pase hacia atrás y el pase hacia adelante y hacia atrás. En particular, el modelo AC-BiLSTM propuesto se implementa con 10 épocas. Tablas 3 para 8 mostrar la matriz de confusión del modelo propuesto tanto para el entrenamiento y la validación como para el entrenamiento y las pruebas. El desempeño del modelo propuesto se muestra en Tablas 9 para 11 para el pase hacia adelante, el pase hacia atrás y los pases hacia adelante y hacia atrás. Estas tablas indican la precisión (P), recuperación (R), puntuación F1 (F1),

puntajes de micropromedio (MIC), macropromedio (MAC) y promedio ponderado (WA). **Cuadro 12** muestra la precisión media de entrenamiento, validación y prueba. La curva de aprendizaje del modelo propuesto se muestra en **Figura 2**, **Fig. 3**, y **Figura 4**. Hemos comparado nuestro modelo AC-BiLSTM propuesto con varios modelos existentes que han utilizado el mismo conjunto de datos y las mismas categorías de etiquetas como se muestra en **Cuadro 13**. En esta tabla, Rashkin et al. [10] utilizó instancias etiquetadas de los sitios PolitiFact y PunditFact. Roy y col. [12] han mostrado resultados más altos. Han utilizado el conjunto de datos de referencia LIAR con un número reducido de instancias. Además, Yunfei Long [13] experimentó con un número limitado de parámetros del conjunto de datos LIAR. Por lo tanto, estimamos que el modelo propuesto se comporta comparativamente mejor que los resultados del estado de la técnica.

5. Conclusión

En este artículo, propusimos una red convolucional de memoria a largo plazo a largo plazo basada en la atención para detectar automáticamente noticias falsas utilizando el conjunto de datos LIAR. Este conjunto de datos proporcionó una breve declaración auténtica y original de diferentes hablantes con diferentes contextos. Por lo tanto, influyó en el desarrollo de un método de detección de noticias falsas en una amplia cobertura. Primero, preprocesamos el conjunto de datos LIAR utilizando la conversión de mayúsculas y minúsculas, eliminando símbolos y puntuación, y técnicas de tokenización. En segundo lugar, generamos vectores de palabras para los datos de entrada dados utilizando la incrustación de palabras previamente entrenadas por GloVe. Finalmente, el modelo AC-BiLSTM propuesto se emplea en estos vectores de palabras para predecir noticias falsas en un entorno de clases múltiples. Específicamente, el modelo propuesto captura local, global, y el significado temporal de la oración usando C-BiLSTM y luego el mecanismo de atención ayuda a memorizar una secuencia de entrada larga. Por lo tanto, nuestro modelo híbrido propuesto logra una mayor precisión (35,1%) y una puntuación micro F1 (39%) que otros modelos existentes. Sin embargo, el método de detección de noticias falsas sigue siendo un problema abierto en términos de datos, modelo, función y aplicación. En trabajos futuros, pretendemos predecir noticias falsas según el género y el grupo de edad utilizando datos de audio, video, imágenes y texto. Además, los modelos basados en transformadores se pueden utilizar para predecir noticias falsas sobre esta fusión de información. Además, la identificación de elementos clave en la difusión de noticias falsas utilizando redes neuronales gráficas puede ser otra posible dirección para futuras investigaciones. 1%) y micro F1-score (39%) que otros modelos existentes. Sin embargo, el método de detección de noticias falsas sigue siendo un problema abierto en términos de datos, modelo, función y aplicación. En trabajos futuros, pretendemos predecir noticias falsas según el género y el grupo de edad utilizando datos de audio, video, imágenes y texto. Además, los modelos basados en transformadores se pueden utilizar para predecir noticias falsas sobre esta fusión de información. Además, la identificación de elementos clave en la difusión de noticias falsas utilizando redes neuronales gráficas puede ser otra posible dirección para futuras investigaciones. 1%) y micro F1-score (39%) que otros modelos existentes. Sin embargo, el método de detección de noticias falsas sigue siendo un problema abierto en términos de datos, modelo, función y aplicación. En trabajos futuros, pretendemos predecir noticias falsas según el género y el grupo de edad utilizando datos de audio, video, imágenes y texto. Además, los modelos basados en transformadores se pueden utilizar para predecir noticias falsas sobre esta fusión de información. Además, la identificación de elementos clave en la difusión de noticias falsas utilizando redes neuronales gráficas puede ser otra posible dirección para futuras investigaciones. y datos de texto. Además, los modelos basados en transformadores se pueden utilizar para predecir noticias falsas sobre esta fusión de información. Además, la identificación de elementos clave en la difusión de noticias falsas utilizando redes neuronales gráficas puede ser otra posible dirección para futuras investigaciones, y datos de texto. Además, los modelos basados en transformadores se pueden utilizar para predecir noticias falsas sobre esta fusión de información. Además, la identificación de elementos clave en la difusión de noticias falsas utilizando redes

Declaración de contribución de autoría de CRediT

Tina Esther Trueman: Conceptualización, Metodología, Redacción - borrador original, Redacción - revisión y edición, Supervisión.
Ashok Kumar J .: Metodología, Software, Validación, Visualización.
Narayanasamy P .: Supervisión. **Vidya J .:** Software, Validación, Investigación, Visualización.

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.

Referencias

- [1] Zeynep Tufekci, Es la edad de oro (que envenena la democracia) de la libertad de expresión. Cableado <https://www.wired.com/story/free-speech-issue-tech-turmoil-nueva-censura/>. Publicado: 16.01.2018, (Consultado el 2 de marzo de 2020).
- [2] David Leonhardt, Stuart A. Thompson, Las mentiras de Trump, 14, The New York Times, 2017.
- [3] Nolan Higdon, The Anatomy of Fake News: A Critical News Literacy Education, University of California Press, 2020.
- [4] Yann LeCun, Yoshua Bengio, Geoffrey Hinton, Aprendizaje profundo, Nature 521 (7553) (2015) 436–444.

- [5] Tomas Mikolov, Ilya Sutskever, Kai. Chen, Greg S. Corrado, Jeff Dean, Representaciones distribuidas de palabras y frases y su composicionalidad, en: *Advances in Neural Information Processing Systems*, 2013, págs. 3111–3119.
- [6] William Yang Wang, Liar liar pants on fire: Un nuevo conjunto de datos de referencia para la detección de noticias falsas, 2017, preprint arXiv arXiv: 1705.00648.
- [7] Sherry Girgis, Eslam Amer, Mahmoud Gadallah, Algoritmos de aprendizaje profundo para la detección de noticias falsas en texto en línea, en: 2018 13th International Conference on Computer Engineering and Systems (ICCES), IEEE, 2018, págs. 93–97.
- [8] Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio, traducción automática neuronal mediante el aprendizaje conjunto de alinear y traducir, 2014, preimpresión arXiv arXiv: 1409.0473.
- [9] Alex Oliveri, Shaban Shabani, Maria Sokhn, Philippe Cudré-Mauroux, Creación de funciones genéricas de tareas para la detección de noticias falsas, en: Actas de la 52a Conferencia Internacional de Ciencias de Sistemas de Hawái, 2019.
- [10] Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, Yejin Choi, Truth of varying shades: Analyzing language in fake news and policy fact-testing, en: Actas de la conferencia de 2017 sobre métodos empíricos en el procesamiento del lenguaje natural, 2017, págs. 2931–2937.
- [11] Junaed Younus Khan, Md Khondaker, Tawkat Islam, Anindya Iqbal, Sadia Afroz, Un estudio de referencia sobre métodos de aprendizaje automático para la detección de noticias falsas, 2019, preprint arXiv arXiv: 1905.04749.
- [12] Arjun Roy, Kingshuk Basak, Asif Ekbal, Pushpak Bhattacharyya, Un marco de conjunto profundo para la clasificación y clasificación de noticias falsas, 2018, preimpresión de arXiv arXiv: 1811.04670.
- [13] Yunfei Long, Detección de noticias falsas a través de perfiles de oradores con múltiples perspectivas, Asociación de Lingüística Computacional, 2017.
- [14] <https://www.politifact.com/>.
- [15] Jeffrey Pennington, Richard Socher, Christopher D. Manning, Glove: Vectores globales para la representación de palabras, en: Actas de la conferencia de 2014 sobre métodos empíricos en el procesamiento del lenguaje natural (EMNLP), 2014, págs. 1532–1543.
- [16] Yoon Kim, Redes neuronales convolucionales para la clasificación de oraciones, 2014, preprint arXiv arXiv: 1408.5882.
- [17] Pierre Sermanet, Soumith Chintala, Yann LeCun, Redes neuronales convolucionales aplicadas a la clasificación de dígitos de números de casa, en: Actas de la 21ª Conferencia Internacional sobre Reconocimiento de Patrones (ICPR2012), IEEE, 2012, págs. 3288–3291.
- [18] Hussein Mouzannar, Yara Rizk, Mariette Awad, Identificación de daños en publicaciones de redes sociales mediante aprendizaje profundo multimodal, en: ISCRAM, 2018.
- [19] Botao Zhong, Xuejiao Xing, Peter Love, Xu Wang, Hambin Luo, Red neuronal convolucional: clasificación de problemas de calidad de construcción basada en el aprendizaje profundo, *Adv. Ing. Informar.* 40 (2019) 46–57.
- [20] Yuanzhi Li, Yang Yuan, Análisis de convergencia de redes neuronales de dos capas con activación relu, en: Avances en los sistemas de procesamiento de información neuronal, 2017, págs. 597–607.
- [21] Yanming Huang, Y. Jiang, Touhidul Hasan, Q. Jiang, Chao Li, Modelo BiLSTM de un tema para la clasificación de sentimientos, en: Actas de la 2da Conferencia Internacional sobre Innovación en Inteligencia Artificial, 2018, págs. 143–147.
- [22] Jay Vasudevan, Barret Zoph, Jonathon Shlens, Quoc V. Le, Búsqueda de arquitectura neuronal para redes neuronales convolucionales. Patente de EE. UU. Núm. 10,521,729, 2019, 31 dic.
- [23] Sepp Hochreiter, Jürgen Schmidhuber, Memoria a corto plazo, *Computación neuronal*. 9 (8) (1997) 1735–1780.
- [24] Zhiyong Cui, Ruimin Ke, Ziyuan Pu, Yinhai Wang, Red neuronal recurrente LSTM bidireccional y unidireccional profunda para la predicción de la velocidad del tráfico en toda la red, 2018, preprint arXiv arXiv: 1801.02143.
- [25] Shlok Gilda, Aviso de violación de los principios de publicación de IEEE: evaluación de algoritmos de aprendizaje automático para la detección de noticias falsas, en: IEEE 15th Student Conference on Research and Development (SCoReD) de 2017, IEEE, 2017, págs. 110–115.
- [26] Jinlong Sun, Wenjuan Shi, Zhutian Yang, Jie Yang, Guan Gui, modelado de comportamiento y linealización de amplificadores de potencia de RF de banda ancha que utilizan redes BiLSTM para sistemas inalámbricos 5G, *IEEE Trans. Veh. Technol.* 68 (11) (2019) 10348–10356.
- [27] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, Eduard Hovy, redes de atención jerárquica para la clasificación de documentos. en: Actas de la conferencia de 2016 del capítulo norteamericano de la asociación de lingüística computacional: tecnologías del lenguaje humano, 2016, págs. 1480–1489.
- [28] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu. Blondel, et al., Scikit-learn: aprendizaje automático en python, *J. Mach. Aprender. Res.* 12 (2011) 2825–2830.
- [29] Tariq Alhindi, Savvas Petridis, Smaranda Muresan, ¿Dónde está su evidencia? Mejorando la verificación de hechos mediante el modelado de justificación, en: Actas del primer taller sobre extracción y verificación de hechos (FEVER), 2018, págs. 85–90.
- [30] Margherita Grandini, Enrico Bagli, Giorgio Visani, Métricas para la clasificación de clases múltiples: una descripción general, 2020, preprint arXiv arXiv: 2008.05756.