

Sentimental LIAR: Corpus Extendido y Modelos de Deep Learning para Clasificación de Afirmaciones Falsas

Procesamiento del Lenguaje Natural

Esteban Reyes Saldaña

Centro de Investigación en Matemáticas

26 de mayo de 2021

Introducción

- Flujo de información de redes sociales ha facilitado el acceso a la información disponible.
- Acceso a internet permite a los usuarios no sólo el consumo si no también la contribución a la información a cada usuario.
- Los beneficios anteriores vienen con el costo de desconocer la veracidad de la información.

Problemas

- ① Campañas de información falsa.
- ② Los usuarios de manera intencional o no, consumen noticias faltas y también la comparten con sus contactos.

Definición

El **fenómeno de noticias falsas** se refiere a la rediodifusión de información que es intencionalmente y verificablemente falsa.

Ejemplos de FakeNews

2016 Elecciones Presidenciales de US

 PLAN YOUR VACCINE COVID-19 POLITICS U.S. NEWS OPINION WORLD BUSINESS WATCH NOW

WORLD

Fake News: How a Partying Macedonian Teen Earns Thousands Publishing Lies

Fake news ran wild during the presidential election – and even though the contest is over, the gold rush continues for dozens of enterprising teens in a remote Macedonian town. Their main cash cow: president-elect Donald Trump. “Nothing can beat Trump’s supporters when it comes to social media engagement,” says 18-year-old Dimitri, who claims to have made \$60,000 in the past six months from the fake news phenomenon. NBC News takes you inside the small town of Veles, where the teens spend big on cars, clubbing and real estate as they cash in on an empire of lies that shook a presidential election.

2016 Genocidio en Myanmar

The New York Times

SUBSCRIBE NOW LOG IN

Deep in Myanmar | What We Know | Visual Investigation | A Small Town Suffers | The Military's Brutality | Democratic Hopes Unraveled | Return to 'Bad Old Days'

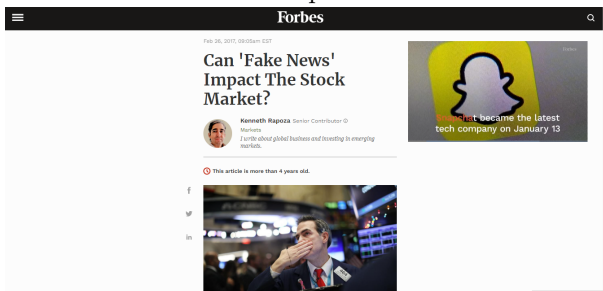
Facebook Admits It Was Used to Incite Violence in Myanmar



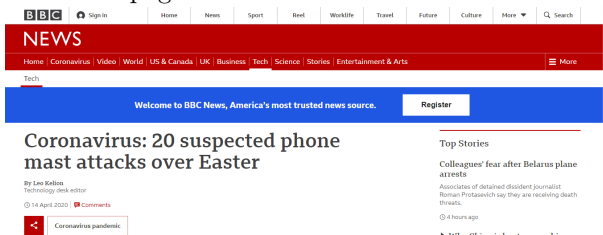


Ejemplos de FakeNews

2017 Barack Obama fue Herido en una Explosión



2020 5G es la causa de Propagación de SARS-COV19



Algunos de los efectos negativos de la exposición de Usuarios a las Noticias Falsas son

- 1 actitudes de ineficacia.
- 2 alienación.
- 3 confiar en propaganda falsa.
- 4 cinismo a personajes políticos o comunidades.
- 5 generación de eventos violentos.

Revisión de la Literatura

Definición

Clasificación de noticias falsas se refiere al proceso de determinar cuando una noticia contiene información falsa/desinformativa o no.

Enfoque Tradicional

- hecha por expertos y periodistas.
- problema: grandes volúmenes de información.

Enfoque Reciente

- automatizar el proceso de clasificación.
- propuesta: usar aprendizaje máquina y NLP para noticias y publicidades.

Algunas definiciones

Definición

Noticia falsa. Se refiere a información fabricada que imita el contenido de los medios de comunicación en forma pero no en el proceso organizativo e intencional.



Hipótesis

Hipótesis de Underutsch

La diferencia entre las noticias falsas y verdaderas está en términos de estilo de escritura (autenticidad, intención)

Teoría de los Cuatro Factores

Las afirmaciones verdaderas falsas y verdaderas son expresadas con distintas emociones y sentimientos.

Hipótesis → emociones y sentimientos ayudan a distinguir afirmaciones verdaderas y falsas.

Bases de Datos

Nombre	Tipo	Tamaño	Etiquetas	Anotadores
FEVER	afirmaciones cortas	185,445	3	<i>personal entrenado</i>
PHEME	hilos de twitter	330	2	<i>periodistas editores</i>
LIAR	afirmaciones cortas	12,836	6	<i>periodistas</i>

$$\text{labels} \begin{cases} \text{pants-fire} & \text{half-true} \\ \text{false} & \text{mostly-true} \\ \text{barely-true} \end{cases} \quad (1)$$

Statement	McCain opposed a requirement that the government buy American-made motorcycles. And he said all buy-American provisions were quote 'disgraceful.'
ID	620
LABEL	True
Speaker	Barack Obama
Speaker Job	President
State Info	Illinois
Party Affiliation	democrat
counts	70, 71, 160, 163, 9
Context	a radio ad

Deep Learning

Autor	Año	meta-data	Modelo	Accuracy
Wang	2017	+speaker + All	CNN	0.274
Long	2017	+All	LSTM	0.415
Kirilin	2018	+All + sp2c	LSTM	0.457

+All significa que se incluyó metadata de LIAR.

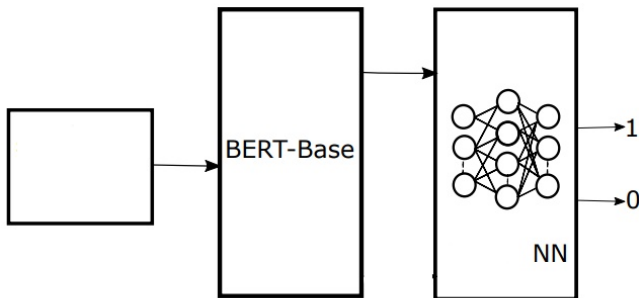
BERT

- 1 Modelo del estado del arte que usa transformers para aprender las relaciones contextuales del texto dado de manera bidireccional.
- 2 Permite ajustes finos para otras tareas a nivel de token.

$$\text{Etapas} \rightarrow \begin{cases} MLM \\ NSP \end{cases}$$

Idea general

Hipótesis : las afirmaciones falsas están escritas en un estilo exagerado y con emociones fuertes.

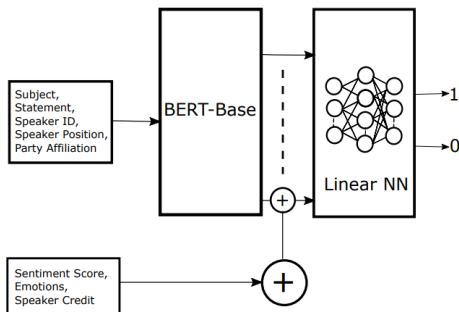


Sentimental-LIAR

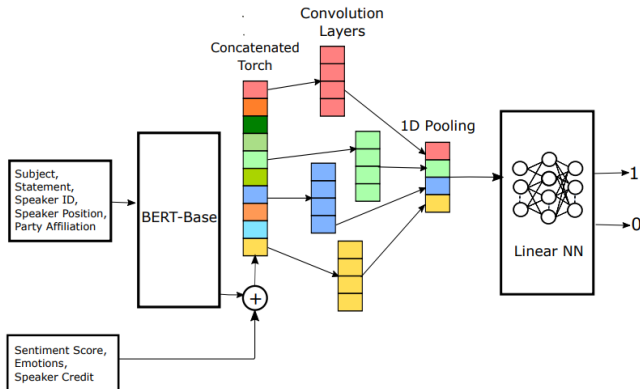
Cuadro: Ejemplo de Sentimental LIAR

TEXT	statement	McCain opposed a requirement that the government buy American-made motorcycles. And he said all buy-American provisions were quote 'disgraceful.'
	subject	federal-budget
	speaker_job	President
	state_info	Illinois
	party_affiliation	democrat
	sentiment	NEGATIVE
EMO	anger	0.1353
	disgust	0.8253
	sad	0.1419
	fear	0.0157
	joy	0.0236
SPC	barely_true_counts	70
	false_counts	71
	half_true_counts	160
	mostly_true_counts	163
	pants_on_fire_counts	9
SEN	sentiment_score	-0.7

BERT + FNN



BERT + CNN



Parámetros Generales

El dataset fue dividido en 80 % entrenamiento, 10 % validación y 10 % pruebas.

parámetro	atributos
batch_size	8
learning_rate	$1e^{-5}$
Dropout (BERT-Base)	0.3
Loss_function	binary_cross_entropy
Optimizer	Adam

Resultados

BERT + FFNN

Experimento	Acc	<i>F1Score Macro</i>
$TEXT \rightarrow [BB]$ $BB_OP \rightarrow [NN]$	0.6882	0.5842
$TEXT + EMO \rightarrow [BB]$ $BB_OP \rightarrow [NN]$	0.6773	0.6352
$TEXT + EMO +$ $SPC \rightarrow [BB]$ $BB_OP \rightarrow [NN]$	0.6720	0.6720
$TEXT + EMO +$ $SPC + SEN \rightarrow [BB]$ $BB_OP \rightarrow [NN]$	0.6720	0.4097
$TEXT \rightarrow [BB]$ $BB_OP + EMO +$ $SPC + SEN \rightarrow [NN]$	0.6937	0.57234

BERT + CNN

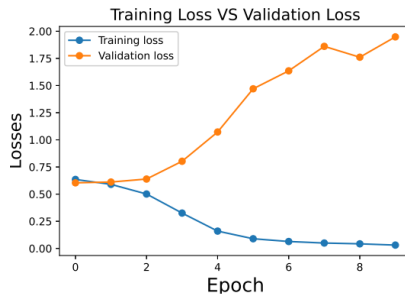
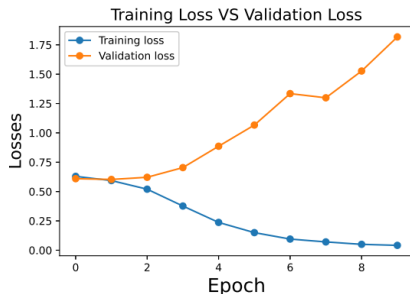
Experimento	Acc	<i>F1Score Macro</i>
$TEXT \rightarrow [BB]$ $BB_OP \rightarrow [CNN]$	0.6882	0.5308
$TEXT + EMO +$ $SPC \rightarrow [BB]$ $BB_OP \rightarrow [CNN]$	0.5546	0.55641
$TEXT \rightarrow [BB]$ $BB_OP + EMO \rightarrow [CNN]$	0.6554	0.608
$TEXT + SPC \rightarrow [BB]$ $BB_OP + EMO +$ $+ SPC \rightarrow [CNN]$	0.6890	.6890
$TEXT \rightarrow [BB]$ $BB_OP + EMO +$ $SPC \rightarrow [CNN]$	0.7000	0.6370
$TEXT \rightarrow [BB]$ $BB_OP + EMO +$ $SPC + SEN \rightarrow [CNN]$	0.6370	0.6430

Conclusiones

- 1 Se extendió el dataset LIAR agregando emociones usando la API de IBM NLP y un puntaje de sentimiento usando la API de Google NLP.
- 2 Tanto para los experimentos realizados con FFNN como con CNN se incrementó el accuracy del estado del arte para el dataset LIAR.
- 3 El mejor modelo fue en el que se le dio TEXT directamente a BERT y luego se concatenó con el resto de los atributos para entrenar la parte de CNN.
- 4 El modelo tuvo un mejor resultado al usar los atributos EMO Y SEN, justo como lo decía la hipótesis de Undeutsch.

Observaciones

- ④ Al revisar las gráficas de las pérdidas, hubo overfitting a partir de la segunda época



Referencias



BURDEN, R. L., AND FAIRES, J. D.

Numerical Analysis, fourth ed.

The Prindle, Weber and Schmidt Series in Mathematics. PWS-Kent Publishing Company, Boston, 1989.



ERWIN, AZRIANSYAH, M., HARTUTI, N., FACHRURROZI, M., AND TAMA, B. A.

A study about principle component analysis and eigenface for facial extraction.

Journal of Physics: Conference Series 1196 (mar 2019), 012010.



JOHNSON, R., AND WICHERN, D.

Applied multivariate statistical analysis, 5. ed ed.

Prentice Hall, Upper Saddle River, NJ, 2002.



PAUL, L., AND SUMAN, A.

Face recognition using principal component analysis method.

International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) 1 (11 2012), 135–139.