EMOTION COGNIZANCE IMPROVES HEALTH FAKE NEWS IDENTIFICATION



Anoop K, University of Calicut Kerala Deepak P, Queen's University Belfast Lajish V L, University of Calicut Kerala







FAKE NEWS Intentionally False

Deliberate injection of misinformation and spread via traditional news or online social media

Fake news is in every domain of news including politics, health, science, business





'INFODEMIC' a new term coined by WHO

'INFODEMIOLOGY' is defined as the science behind managing infodemics

'Fake news threatens our democracy. Fake medical news threatens our lives' - NYTimes





HEALTH FAKE NEWS (HFN) & EMOTION

- > Health fake news originates in online websites
- > The core point of the news is conveyed through long, nuanced textual narratives
- > Abundance of anecdotes, conceivably to appeal to the readers' own experiences or self-conscious emotions
- Scientific claims or content exaggerated or distilled either knowingly or maliciously (to attract eyeballs)
- Empathic Media the use of personally and emotionally targeted news produced by journalism



RELATED WORKS - FAKE NEWS DETECTION

Major research efforts has focused on political domain within microblogging

Work	Task Setting	Target Domain	Features Used		
			Content	Network	Temporal
Kwon et al., 2013	Supervised	Twitter	✓	✓	1
Zubiaga et al., 2017	Supervised	Twitter	✓	~	1
Qazvinian et al., 2011	Supervised	Twitter	✓	~	1
Wu and Liu, 2018	Supervised	Twitter	✓	~	1
Ma et al., 2016	Supervised	Twitter	✓	×	1
Zhao et al., 2015	Supervised	Twitter	✓	×	1
Ma et al., 2017	Supervised	Twitter	✓	×	1
Guo et al., 2019	Supervised	Weibo	✓	~	1
Zhang et al., 2017	Unsupervised	Weibo	✓	×	1
Zhang et al., 2016	Unsupervised	Weibo	✓	×	\checkmark





RELATED WORKS - FAKE NEWS & EMOTIONS

> [Bhutani et al., 2019]: Make use of sentiment scores.

- [Guo et al., 2019]: Works within microblogging platforms and extensively exploit the usage of publisher emotions and social emotions.
- [Paschen et al., 2019]: Reports significantly more negative emotions in the titles of the fake news





EXAMPLE: HFN HEADLINE AND EXCERPT

Revolutionary juice that can burn stomach fat while sleeping

Having excess belly fat poses a serious threat to your health. Fat around the midsection is a strong risk factor for heart disease, type 2 diabetes, and even some types of cancers. Pineapple-celery duo is an ideal choice for those wanting to shed the fat deposits around the stomach area due to the presence of enzymes that stimulate the fat burning hormones. All you need to do is drink this incredible burn-fat sleeping drink and refrain from eating too much sugar and starch foods during the day.



Association for Computing Machinery



EMPIRICAL SETTINGS



- Our focus: Exploit the differences in emotional content of the fake and real news through word-addition-based text transformations.
 - > The identification of the nature of differences is outside the scope of the current study

acm

Association for



Inputs:

- \succ News document: D
- > Emotion lexicon: L (**NRC Intensity Emotion Lexicon** [Saif M Mohammad, 2017])
 - \circ 3-tuples, [*w*, *e*, *s*]
 - *W*: word
 - $\circ e$: emotion
 - $\circ s \in [0, 1]$: intensity quantified
 - Example: [death, sadness, 0.915], [death, fear, 0.843] & [death, anger, 0.574]
- \succ Threshold: au

Output:

 \succ Emotionized document D'



Association for Computing Machinery

 $D, \mathcal{L} \xrightarrow{Emotionization} D'$





Association for Computing Machinery

















Association for Computing Machinery























Emotionized Health Fake News Excerpt (added emotion labels in bold)

Revolutionary juice that can burn stomach fat while sleeping

Having excess belly fat poses a serious threat **anger** to your health. Fat around the midsection is a strong risk **fear** factor for heart disease **fear**, type 2 diabetes, and even some types of cancers **sadness**. Pineapple-celery duo is an ideal choice for those wanting to shed the fat deposits around the stomach area due to the presence of enzymes that stimulate the fat burning hormones. All you need to do is drink this incredible burn-fat sleeping drink and refrain from eating too much sugar and starch foods **joy** during the day.





EMPIRICAL STUDY - DATASET

We curated a new dataset **HWB**, of fake and legitimate news articles within the topic of health and well being

- 500 Real news CNN, NYTimes, New Indian Express, etc.
- 500 Fake news BeforeltsNews, Nephef, MadWorldNews, etc.

Dataset	Class	Total Number of Documents in the Class	Average Words per Document	Average Sentences per Document	Total Number of Words
Health and Well Being (HWB)	Real	500	724	31	362117
	Fake	500	578	28	289477

https://dcs.uoc.ac.in/cida/resources/hwb.html





EMPIRICAL STUDY - SUPERVISED SETTING

Conventional Classifiers:

- Naive Bayes (NB)
- k-Nearest Neighbour (KNN)
- Support Vector Machine (SVM)
- Random Forests (RF)
- Decision Tree (DT)
- > AdaBoost (AB)

Neural Networks:

- Long Short Term Memory (LSTM)
- Convolutional neural network (CNN) [Yoon Kim, 2014]





EMPIRICAL STUDY - SUPERVISED SETTING

Input Data:

 $> D = \{\dots, D, \dots\}$: corpus of all news articles

 $> D' = \{\dots, D', \dots\}$: corresponding emotionized corpus

 \succ DBOW doc2vec used to build vectors over two corpus of vectors, V and V'

 $> d = \begin{cases} \text{dimensionality of the document vector, for the conventional classifiers} \\ \text{dimensionality of the word vectors, for the CNN and LSTM classifiers} \end{cases}$

Actual Labels:

 \geq Each document is labeled as either fake or not (0/1)

Evaluation Measure:

> Accuracy: Acc(D) & Acc(D')





SUPERVISED SETTING - RESULTS

Mathad	$Acc(\mathbf{D})$	$Acc(\mathcal{D'})$						
Methou	Att(D)	$\tau = 0.0$	$\tau = 0.2$	$\tau = 0.4$	$\tau = 0.6$	$\tau = 0.8$		
	Classification Results for $d = 100$							
NB	77.0	78.0	78.0	78.5	79.0	77.5		
KNN	75.0	75.0	75.5	76.0	92.5	75.0		
SVM	50.0	65.0	75.0	75.0	90.0	70.0		
RF	63.0	71.0	70.0	72.0	84.0	80.5		
DT	68.0	69.0	70.0	78.0	94.0	78.5		
AB	55.0	57.0	70.0	71.0	96.5	82.5		
CNN	87.0	88.0	90.0	88.0	91.0	88.0		
LSTM	90.5	90.0	91.0	91.0	92.0	92.0		
	Cla	assificatio	n Results	for $d = 30$)0			
NB	77.0	80.0	81.0	79.0	83.0	78.0		
KNN	72.0	74.0	75.0	76.0	91.0	74.5		
SVM	60.0	67.0	72.0	74.0	89.0	72.0		
RF	65.0	70.0	73.0	71.5	82.0	75.0		
DT	60.0	65.0	73.0	78.0	90.5	75.0		
AB	55.0	55.0	72.0	81.0	94.5	75.0		
CNN	91.2	91.0	92.7	92.0	92.0	91.0		
LSTM	90.0	90.2	90.0	90.2	90.7	90.0		



Association for Computing Machinery



EMPIRICAL STUDY - UNSUPERVISED SETTING

Clustering methods: K-Means & DBSCAN

K-Means: Top-down approach

 $\Box k$ = Desired number of output clusters

DBSCAN: bottom-up approach

 $\Box ms =$ Minimum samples to form a cluster

Evaluation Measure: Clustering Purity: Pur(D) & Pur(D')





UNSUPERVISED SETTING - RESULTS

	$D_{ur}(\mathbf{O})$	$Pur(\mathcal{D}')$						
	Fur(D)	$\tau = 0.0$	$\tau = 0.2$	$\tau = 0.4$	$\tau = 0.6$	$\tau = 0.8$		
k	K-Means Clustering Results for $d = 100$							
2	52.3	52.4	52.3	52.3	56.1	52.9		
4	78.1	78.0	78.6	79.3	81.6	79.3		
7	85.0	85.7	85.2	85.1	86.9	85.6		
10	85.3	85.1	85.1	85.1	87.7	85.7		
15	85.2	85.3	85.1	85.1	87.8	85.8		
20	85.2	85.2	85.0	85.1	88.7	85.7		
k	K-Means Clustering Results for $d = 300$							
2	51.3	52.0	52.0	52.0	55.5	52.0		
4	77.1	77.8	78.1	78.9	81.5	78.5		
7	84.0	84.0	85.0	84.9	86.9	84.6		
10	85.0	85.0	85.0	85.0	87.1	85.1		
15	85.1	85.3	85.1	85.1	87.5	85.2		
20	85.0	85.2	85.0	85.0	88.0	85.0		



Association for Computing Machinery

K-Means



UNSUPERVISED SETTING - RESULTS

	$D_{uv}(\mathbf{O})$	$Pur(\mathcal{D'})$						
	Pur(D)	$\tau = 0.0$	$\tau = 0.2$	$\tau = 0.4$	$\tau = 0.6$	$\tau = 0.8$		
ms	DBSCAN Clustering Results for $d = 100$							
20	61.0	62.0	62.0	62.0	65.0	61.9		
40	62.7	65.5	64.5	58.1	66.5	65.0		
60	71.6	72.1	72.0	72.5	72.5	72.5		
80	85.1	85.0	85.1	85.6	86.0	85.6		
100	84.5	84.1	84.8	84.7	86.0	84.0		
ms	DBSCAN Clustering Results for $d = 300$							
20	61.0	61.5	61.0	61.0	63.5	62.0		
40	63.5	66.3	66.5	66.9	67.0	65.5		
60	67.5	70.1	70.5	71.0	71.5	70.0		
80	78.0	81.0	81.9	82.0	82.5	80.8		
100	75.5	80.0	80.0	80.0	80.5	80.0		



Association for Computing Machinery

DBSCAN

EMOTIONIZATION & COVID-19 FAKE NEWS

Do not consent to nose swab testing!

Avoid **fear** the Covid-19 test at all costs. These swabs may be (and probably are) contaminated **fear** with something dangerous **fear**, like viruses or something we don't understand. People should be just as concerned **fear** with the swab as they are about the vaccine. I was wondering why the PCR test for COVID-19 had to be so far back and it got me thinking...how far does it go? So I did some research and found these two pictures and overlapped them. The suprising **joy** evidence was shocking **fear**! The blood **fear** brain barrier **anger** is exactly where the swab test has to be placed.

An example of Emotionized COVID-19 Fake News (added emotion labels in bold)



Association for Computing Machinery

FUTURE WORK



- Emotion-aware end-to-end methods by blending the emotion cues with collective behavior heuristics [Gangireddy et al., 2020]
- Considering the use of lexicons learned from data itself [Bandhakavi et al., 2014]
- > Exploring the usage of the affective content of responses to social media posts



SUMMARY



- > Considered the utility of the affective character for the task of HFN detection
- Procured a new dataset HWB for the task
- > Proposed a simple method to emotionize text using external emotion intensity lexicons
- > Empirically evaluated the representations and noteworthy gains were observed
- > Firmly establishes the utility of emotion information in improving HFN detection







CONTACT: ANOOPK_DCS@UOC.AC.IN



Association for Computing Machinery