An Emotion-Based Multi-Task Approach to Fake News Detection (Student Abstract)

Arjun Choudhry[*†], Inder Khatri[*], Minni Jain[*]
Delhi Technological University, New Delhi, India
{arjunchoudhry_2k18it031, inderkhatri_2k18mc043, minnijain}@dtu.ac.in

Abstract

Social media, blogs, and online articles are instant sources of news for internet users globally. But due to their unmoderated nature, a significant percentage of these texts are fake news or rumors. Their deceptive nature and ability to propagate instantly can have an adverse effect on society. In this work, we hypothesize that legitimacy of news has a correlation with its emotion, and propose a multi-task framework predicting both the emotion and legitimacy of news. Experimental results verify that our multi-task models outperform their single-task counterparts in terms of accuracy.

Introduction

In recent years, we have witnessed a substantial increase in the usage of social media, news services and blogs, leading to an exponential increase in the spread of *fake news* and *rumors*. This calls into question the credibility of social media and the web as a source of information. To this end, significant work has been done to tackle the problem of *fake news detection*. Recently, Pérez-Rosas et al. (2018) presented two novel datasets, *FakeNews AMT* and *Celeb*, for fake news detection across multiple domains, and used hand-crafted linguistic features and an SVM model for fake news detection. Saikh et al. (2020) treated fake news detection as a text classification task, and presented two deep learning models for fake news detection on FakeNews AMT and Celeb datasets. In this work, we hypothesize a relation between the legitimacy of news and its emotion and propose an emotion-based multi-task approach for fake news and rumor detection.

Methodology

Datasets & Pre-processing

We use PHEME 9 (Zubiaga, Liakata, and Procter 2016), FakeNews AMT and Celeb datasets for fake news detection. During pre-processing, we convert the text to lower case, decontract verbs forms (eg. *"I'll"* to *"I will"*) and remove all punctuation marks.

[*] All the authors contributed equally, and wish that they be regarded as joint First Authors.

[†] Mobile Number: +91 7768955600

Emotion Classes & Annotating the Datasets

Incorporating emotion classification as an auxiliary task in our multi-task framework poses two challenges: annotating the datasets with emotion classes, and choosing a set of basic emotions. There are two widely accepted theories on *basic* emotion classes, proposed by Plutchik (1982) and Ekman (1992) respectively. Based on his *Ten Postulates*, Plutchik designed a wheel with 8 basic emotions (*Joy*, *Surprise*, *Trust*, *Anger*, *Anticipation*, *Sadness*, *Disgust*, *Fear*). During his cross-cultural study, Ekman inferred that there are 6 basic emotions (*Joy*, *Surprise*, *Anger*, *Sadness*, *Disgust*, *Fear*), each considered a discrete category. Due to unavailability of fake news datasets annotated with emotion classes, we use the *Unison model* (Colneric and Demsar 2018) to generate both Plutchik and Ekman emotion tags for our datasets, and compare the performance of various classifiers in multi-task settings on both emotion sets.

Intuition behind Emotion for Fake News Detection

Recent works in fake news and rumor detection have shown the efficacy of using sentiment analysis and text polarity for fake news and rumor detection. Ajao, Bhowmik, and Shahrzad. (2019) calculated *emoratio* (ratio of negative polarity to positive polarity in text) for PHEME dataset, showing a significant difference in values for rumors and non-rumors. Augmentation of emoratio into the feature matrix showed noticeable improvements to model performance. Yang et al. (2018) showed that real news has higher median values and lower standard deviation for both positive and negative sentiments than fake news. The median values for negative sentiment for fake news were also higher than the median values for positive sentiments. Both indicate more negative sentiments are associated with fake news.

We expand sentiment analysis for fake news detection by treating it as a multi-class emotion classification task. Figure 1 represents 3-dimensional Principal Component Analysis (PCA) graphs for PHEME dataset using embeddings generated from the Unison model for Plutchik emotions, plotted separately for rumors and non-rumors. Non-rumors (Figures 1A and 1C) show better formed clusters and a higher percentage of *Trust* and *Fear* than rumors (Figures 1B and 1D), which show a higher percentage of *Sadness* and *Surprise*. This indicates that a relation exists between legitimacy of rumors and emotions, thereby supporting our hypothesis.

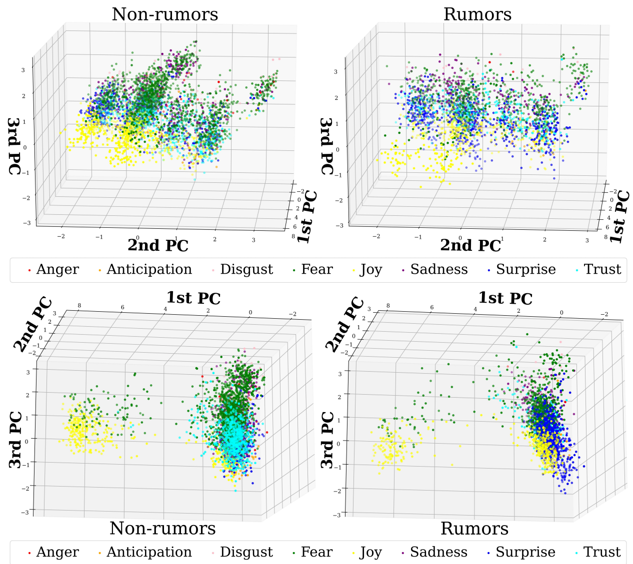

Figure 1: 3-dimensional PCA plots for Rumors and Non-rumors, colored by Plutchik emotions.

Multi-task Learning for Fake News Detection

Multi-task learning (MTL) uses shared representations between primary and auxiliary tasks for better extraction of common features, which can otherwise get ignored. To verify the relation between the legitimacy and emotion of news, we train our models to predict both the legitimacy and emotion for a text, and evaluate them against their single-task (STL) counterparts. We further leverage the domain tags for texts in FakeNews AMT dataset, and evaluate the performance of MTL models predicting both legitimacy and domain of news.

Classification Models

We evaluate the performance of LSTM, CNN-LSTM (CLSTM) and BERT models in both MTL and STL settings. For the PHEME 9 dataset, we also evaluate the performance of CNN, HAN and CapsuleNet models. We further compare the performance of classifiers using our approach with other novel approaches on FakeNews AMT and Celeb datasets.

Results and Discussion

We evaluated the performance of our proposed framework across a number of datasets and deep learning models, using the same train-test split for each dataset across all classifiers. Table 2 illustrates the results of our experiments on the PHEME 9 dataset, while Table 1 compares the results of our proposed approach with other novel models on the FakeNews AMT and Celeb datasets. Some findings observed are:

MTL models outperform their STL counterparts We observed an improvement in the performance of classifiers in MTL settings over STL, verifying our hypothesis about the correlation between legitimacy and emotion of news.

MTL models with Ekman and Plutchik emotions perform comparably We observed comparable performance between MTL models trained on Ekman (*MTL(6)*) and Plutchik (*MTL(8)*) emotions across all datasets, with the exception of LSTM and CNN-LSTM models on Celeb dataset, which performed worse with Plutchik emotions.

MTL models with domain classification as auxiliary task outperform their STL counterparts Classifiers trained to identify both domain and legitimacy of news (*MTL(D)* in Table 1) outperformed their STL counterparts on the FakeNews AMT dataset, achieving accuracy comparable to the emotion-based MTL models. This indicates a deeper relation between legitimacy of news and its domain.

Dataset	Setting	SVM	ELMo	LSTM	CLSTM	BERT
FAMT	STL	0.74	0.833	0.725	0.733	0.816
	MTL(6)	-	-	**0.775**	0.758	**0.875**
	MTL(8)	-	-	0.758	**0.766**	0.866
	MTL(D)	-	-	**0.775**	0.758	0.866
Celeb	STL	0.76	0.790	0.736	0.704	0.816
	MTL(6)	-	-	**0.752**	**0.712**	0.856
	MTL(8)	-	-	0.712	0.696	**0.880**

Table 1: Performance evaluation on FakeNews AMT (FAMT) and Celeb datasets using accuracy. MTL models outperform their STL counterparts. BERT with MTL outperforms the SVM model by Pérez-Rosas et al. (2018) and ELMo model by Saikh et al. (2020).

Setting	CNN	LSTM	CLSTM	CapsNet	HAN	BERT
STL	0.870	0.857	0.860	0.853	0.847	0.859
MTL(6)	0.874	0.880	**0.876**	0.858	**0.867**	**0.884**
MTL(8)	**0.875**	**0.881**	0.872	**0.863**	**0.867**	0.874

Table 2: Performance evaluation on PHEME 9 dataset using accuracy. MTL models outperform their STL counterparts.

References

Ajao, O.; Bhowmik, D.; and Shahrzad., Z. 2019. Sentiment Aware Fake News Detection on Online Social Networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP)*, 2507–2511. Brighton, United Kingdom: IEEE.

Colneric, N.; and Demsar, J. 2018. Emotion Recognition on Twitter: Comparative Study and Training a Unison Model. *IEEE Transactions on Affective Computing*, 11(3): 433–446.

Ekman, P. 1992. An argument for basic emotions. *Cognition & Emotion*, 6: 169–200.

Pérez-Rosas, V.; Kleinberg, B.; Lefevre, A.; and Mihalcea, R. 2018. Automatic Detection of Fake News. In *Proceedings of the 27th International Conference on Computational Linguistics*, 3391–3401. Santa Fe, New Mexico, USA: ACL.

Plutchik, R. 1982. A psychoevolutionary theory of emotions. *Social Science Information*, 21(4-5): 529–553.

Saikh, T.; De, A.; Ekbal, A.; and Bhattacharyya, P. 2020. A Deep Learning Approach for Automatic Detection of Fake News. arXiv:2005.04938.

Yang, Y.; Zheng, L.; Zhang, J.; Cui, Q.; Li, Z.; and Yu, P. S. 2018. TI-CNN: Convolutional Neural Networks for Fake News Detection. arXiv:1806.00749.

Zubiaga, A.; Liakata, M.; and Procter, R. 2016. Learning Reporting Dynamics during Breaking News for Rumour Detection in Social Media. arXiv:1610.07363.