2020

Credibility Analysis of News on Twitter using LSTM: An exploratory study

Piyush Vyas
Dakota State University

Omar F. El-Gayar
Dakota State University

Aug 10th, 12:00 AM

Credibility Analysis of News on Twitter using LSTM: An exploratory study

Piyush Vyas
Dakota State University, piyush.vyas@trojans.dsu.edu

Omar El-Gayar
Dakota State University, omar.el-gayar@dsu.edu

Credibility Analysis of News on Twitter using LSTM: An exploratory study

Emergent Research Forum (ERF)

Piyush Vyas
Dakota State University
piyush.vyas@trojans.dsu.edu

Omar El-Gayar
Dakota State University
omar.el-gayar@dsu.edu

Abstract

The popularity and pervasiveness of social media platforms as mechanisms for the rapid dissemination and propagation of news and the ease by which such information can be created and shared, makes it increasingly important to verify its credibility. This work focuses on the automatic credibility analysis of news on microblogging platforms such as Twitter. Using the publicly available PHEME twitter dataset, we perform classification using the Long Short-Term Memory (LSTM) technique. Our dataset was divided into two parts, 80% for training and validation, and 20% for testing. The preliminary result s show the potential of our proposed model to classify news tweets as credible (or non-credible) based on tagging and textual features.

Keywords

Credibility analysis, Twitter, LSTM, social media analytics, text classification.

Introduction

The Internet has established itself as a primary source of information for most people across the globe. Within the Internet, microblogging platforms like Twitter have gained the attention of users as a mechanism for propagating news at an ever-increasing pace (Alrubaian et al. 2018). With the democratization of the capture and dissemination of news, it has become increasingly important to verify the credibility of the news content shared on such platforms.

In that regard, several sites are now available to validate various news in circulation, e.g., Snopus.com and politifact.com (Popat et al. 2016). However, these sites rely predominantly on manual processes for fact-checking and validating news contents. Such manual processes can become expensive, very quickly given the sheer volume of content. Accordingly, there are several approache s have been proposed that aim to automate the manual processes using machine learning techniques (Alrubaian et al. 2018; Boididou et al. 2018; Castillo et al. 2011).

However, these approaches exhibit a number of limitations. First, most similar studies, such as (Boididou et al. 2018; Castillo et al. 2011), concentrated on handcrafted features from twitter posts, thereby limiting its scalability. Second, other studies in this area (Alrubaian et al. 2018; Boididou et al. 2018; Castillo et al. 2011) focused on traditional machine learning techniques such as Support Vector Machine (SVM), Decision Tree (DT-rank), which are more suited to the classification of structured temporal data rather than small-scale sequential data. Third, researchers who have analyzed news credibility on large -scale contextual data such as news articles (Popat et al. 2016, 2018) have not focused on small-scale data such as tweets.

Accordingly, the objective of this research is to complement prior research by automating the credibility analysis of news on microblogging sites such as Twitter. This is accomplished by utilizing textual and tagging features from tweets and exploring the use of deep learning techniques such as Long Short-Term Memory (LSTM) for classification. Specifically, we aim to take advantage of the LSTM technique due to its potential for learning continuous representations of microblog events for a ssessing the credibility of tweets, i.e., classifying if a particular tweet is credible or not.

From a theoretical perspective, the identification of textual and tagging feature set will help to extend the theory related to credibility analysis. Further, the research explores the potential and limitations of LSTM in analyzing micro-blogs. From a practical perspective, the proposed model provides an additional tool in the perpetual struggle to identify and mitigate the propagation of fake news.

Literature Review

Credibility Analysis of Social Media

Over the past decade, dissemination of fake news on various social media platforms has caught the attention of information system researchers. Existing literature has highlighted the problem of spreading rumors (Ma et al. 2016) on social media and microblogging sites (Qazvinian et al. 2011). Assessment of the trustfulness of fact-checking news articles on blogging websites was also investigated (Popat et al. 2016, 2018). People rely on investigative journalism and their common sense to analyze the content of news to distinguish rumors (Ma et al. 2016). Nowadays, some fact-checking social media platforms such as Snopes.com provide claims about the truthfulness of news articles (Popat et al. 2016). Evidence aware deep learning opens the path for investigating misinformation (Popat et al. 2018). Furthermore, there is an increasing need for joint assessment of credibility and trustworthiness of users, articles, and news sources as users assume the role of citizen journalists. (Mukherjee and Weikum 2015).

The spread of fake news tweets regarding real-world events such as the disappearance of Malaysian Airlines flight MH370 has posed a crucial problem in the past. Users who post fake tweets for providing benefits to a particular event are also responsible for spreading misleading information. The identification of credible information is becoming a challenging task for researchers. Content-based, network-based, and specific memes are the features that can help to identify the user who spreads the rumors across the microblogging platform (Qazvinian et al. 2011). There were also attempts to evaluate the credibility of images and videos that accompany tweets and are presented as evidence (Boididou et al. 2018). Overall, analyzing tweet credibility proves to be a promising yet challenging endeavor (Castillo et al. 2011).

Long Short-Term Memory

Recurrent neural networks (RNN) can facilitate the processing of sequential data and are applied to text or speech recognition (Heinrich et al. 2019.). However, standard RNNs become unable to learn long-term dependencies. To address this issue, LSTM was introduced as a successful architecture to obtain remarkable performance in statistical machine translation (Zhou et al. 2015). The potential of LSTM to learn long-distance temporal dependencies with gradient-based optimization make it perform better in short text classification. This is accomplished by controlling the information of the input sequence using three gate structures: input gate, forget gate, and output gate. These gates allow LSTM to store initial input data units in memory cells, storing forgotten data units in existing memory cells and storing the new output data units in memory cells (Ma et al. 2016; Na 2020).

With respect to social media mining, traditional techniques such as Support Vector Machine (SVM) and Decision Tree (DT-rank) (Alrubaian et al. 2018; Boididou et al. 2018; Castillo et al. 2011) tend to focus on the classification of structured temporal data rather than small-scale sequential data that further limits its scalability (Ma et al. 2016.; Popat et al. 2018). These techniques were often applied on handcrafted features from twitter posts (Boididou et al. 2018; Castillo et al. 2011) and as non-parametric models, they tend to be computationally expensive (Karmiani et al. 2019) and possibly underperform recurrent neural network-based models (Ma et al. 2016) such as LSTM. Further, prior research work (Ma et al. 2016; Popat et al. 2016, 2018) demonstrated the potential performance of RNN to using limited unstructured textual data set. This research is complementing prior research by taking advantage of the LSTM technique due to its potential for learning continuous representations of microblog events for identifying and assessing the credibility of tweets.

An LSTM Model for Credibility Analysis

Data Preprocessing

Data preprocessing is essential to improve the performance of the model and to remove redundancies from the collected data. The data set comprises different language reviews along with shared media content (images and videos). We eliminated those pieces of information which are not expressive. Specifically, we preprocessed the data to remove stop words, punctuations, period signs, white spaces, and line breaks. We implemented case-folding to convert text into lower case. For utilizing the textual and tagging feature, we removed irrelevant data fields such as empty cells, images, and videos.

Feature Extraction

Each tweet has a length restriction. Previously it was 140, and now it has a 280-characters limit. Features such as number of words, number of characters, tweet text are categorized as textual features. Features such as unified resource locators (URLs), hashtags (#), at-sign (@)) are also essential to assess tweet credibility. Such features are categorized as tagging features that help to analyze the source credibility for the contents of a tweet. We used Global vectors (GloVe) (Pennington et al. 2014) for feature extraction, which is an extended version of the word2vec model.

Model Development

Figure 1 represents the proposed model for the automatic credibility analysis of news tweets based on the selected news events provided by the data set. In this work, for automatic classification, we propose the use of LSTM due to its ability to handle variable-length sequential data such as sentences and mitigates some of the restrictions on RNN. Examples of such restrictions include the inability to learn long-distance temporal dependencies with gradient-based optimization (Ma et al. 2016). An extension of the memory unit for information storage over a long period can eliminate this restriction. The traditional recurrent unit's state is overwritten each time, whereas the LSTM unit manages a memory cell at a time. The degree of new memory, logistic sigmoid, forgot gate, and updating new memory is a primary component of LSTM (Kochkina et al. 2018). Furthermore, LSTM showed noticeable results with short sentences (limited characters) like tweets (Kochkina et al. 2018).

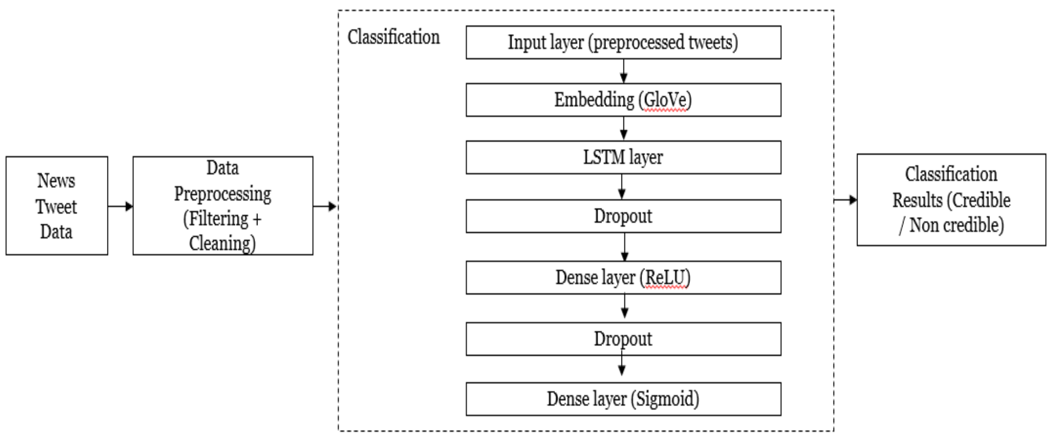

Figure 1. Proposed Auto Credibility Classification Model

Processed tweets are fed to the input layer of the classification model. The embedding layer encodes the input into real-valued vectors using lookup tables. For word representation, we utilized the pre-trained word vector named GloVe. The embedding step performs mapping of tokenized word and vector table. The LSTM layer is an enhancement version of RNN. The dropout layers prevent the model from overfitting by randomly dropping units from the neural network during training. We used the ReLU and Sigmoid for activation (Kochkina et al. 2018; Ma et al. 2016). Both are non-linear functions that capture complex data patterns. We use a sigmoid function in the output layer to predict values of zero (non-credible) or one (credible) in classification.

Evaluation

Data Source

Annotation of the tweet as fake or real is a challenging task. The PHEME is a collection of tweets that contains unbiased categorized data of tweets. It is a publicly available data set that consists of thousands of tweets grouped into five real-world news events such as shooting in Ottawa, the hostage situation in Sydney, and other events (Kochkina et al. 2018). The data is distributed in two root directories: rumor and non-rumor. Data has various fields such as tweet id, text, URLs, retweets, timestamp, username.

Preliminary Results Discussion

Table 1 shows our preliminary results with contextual features and tagging features.

Textual features						Tagging + Textual features					
Training		Validation		Test		Training		Validation		Test	
Accuracy	loss	Accuracy	loss	Accuracy	loss	Accuracy	loss	Accuracy	loss	Accuracy	loss
82%	0.5	81%	0.5	79%	0.5	75%	0.6	74%	0.6	72%	0.6

Table 1. Preliminary Results

Table 2 shows the performance of different methods on textual features of tweets. Overall, the preliminary results are comparable to those obtained using existing methods such as SVM-DSTS (Ma et al. 2015) and DT-Rank (Castillo et al. 2011) with respect to evaluation parameters such as Precision, Recall, F1 score and Validation Accuracy. Precision refers to the percentage of relevant results by our model; recall refers to the percentage of total correctly classified results by our model, and a weighted combination of precision and recall is known as f1 score.

Methods	Precision	Recall	F1 Score	Accuracy
SVM DSTS (Ma et al. 2015)	0.8	0.9	0.8	85%
DT-Rank (Castillo et al. 2011)	0.8	0.8	0.8	86%
LSTM	0.5	0.7	0.6	81%

Table 2. Comparison with other methods.

However, the proposed model is demonstrating a better ability to capture the temporal aspects of the posts and thus potentially outperforming existing models once hyperparameter and optimization tuning are completed. Additional performance improvement plans include the use of loss normalization and layer normalization for LSTM.

Conclusion

Earlier research work relied mostly on the manual assessment using Amazon Mechanical Turk services for credibility analysis of social media content. While there have been attempts to automate the process for analyzing the credibility of tweeter posts, these approaches exhibit a number of limitations. Herein, we propose the use of LSTM deep learning model to address some of these limitations. Specifically, the proposed model has the ability to learn continuous representations of microblog events thereby potentially improving the credibility analysis for microblogs such as Tweeter. From a theoretical perspective, the identification of textual and tagging feature set will help to extend the theory related to credibility analysis. Further, the research explores the potential and limitations of LSTM in analyzing micro-blogs. From a practical perspective, the proposed model provides an additional tool for identification and mitigation of fake news dispersion on Twitter. Specifically, the proposed model (classifier) may be instantiated within a system for identifying fake news as soon as it is created and from the subsequent disclosure of lack of credibility.

Overall, the initial results are somewhat comparable to existing models yet demonstrates that the potential exists for the proposed LSTM model to outperform such models once hyperparameter optimization and tuning, and additional model enhancements are completed. Accordingly, future research will focus on the optimization of the classification performance with the help of evolutionary algorithms such as Genetic and particle swarm optimization. Future research can also focus on issues surrounding the instantiation of the proposed model within a larger system for the dynamic identification and disclosure of tweets lacking credibility.

REFERENCES

Alrubaian, M., Al-Qurishi, M., Hassan, M. M., and Alamri, A. 2018. "A Credibility Analysis System for Assessing Information on Twitter," *IEEE Transactions on Dependable and Secure Computing* (15:4), pp. 661–674. (https://doi.org/10.1109/TDSC.2016.2602338).

Boididou, C., Papadopoulos, S., Zampoglou, M., Apostolidis, L., Papadopoulou, O., and Kompatsiaris, Y. 2018. "Detection and Visualization of Misleading Content on Twitter," *International Journal of Multimedia Information Retrieval* (7:1), pp. 71–86. (https://doi.org/10.1007/s13735-017-0143-x).

Castillo, C., Mendoza, M., and Poblete, B. 2011. "Information Credibility on Twitter," in *Proceedings of the 20th International Conference on World Wide Web - WWW '11*, New York, New York, USA: ACM Press, p. 675. (https://doi.org/10.1145/1963405.1963500).

Heinrich, K., Zschech, P., Skouti, T., Griebenow, J., and Riechert, S. (2019). *Demystifying the Black Box: A Classification Scheme for Interpretation and Visualization of Deep Intelligent Systems*, p. 10.

Karmiani, D., Kazi, R., Nambisan, A., Shah, A., and Kamble, V. 2019. "Comparison of Predictive Algorithms: Backpropagation, SVM, LSTM and Kalman Filter for Stock Market," in *2019 Amity International Conference on Artificial Intelligence (AICAI)*, , February, pp. 228–234. (https://doi.org/10.1109/AICAI.2019.8701258).

Kochkina, E., Liakata, M., and Zubiaga, A. 2018. "All-in-One: Multi-Task Learning for Rumour Verification," in *Proceedings of the 27th International Conference on Computational Linguistics*, Santa Fe, New Mexico, USA: Association for Computational Linguistics, pp. 3402–3413. (https://www.aclweb.org/anthology/C18-1288).

Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B. J., Wong, K.-F., and Cha, M. (2016). "Detecting Rumors from Microblogs with Recurrent Neural Networks."

Ma, J., Gao, W., Wei, Z., Lu, Y., and Wong, K.-F. 2015. *Detect Rumors Using Time Series of Social Context Information on Microblogging Websites*, , October 1. (https://doi.org/10.1145/2806416.2806607).

Mukherjee, S., and Weikum, G. 2015. "Leveraging Joint Interactions for Credibility Analysis in News Communities," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management - CIKM '15*, Melbourne, Australia: ACM Press, pp. 353–362. (https://doi.org/10.1145/2806416.2806537).

Na, R. 2020. "GADM: Manual Fake Review Detection for O2O Commercial Platforms," *Computers & Security* (88), p. 101657. (https://doi.org/10.1016/j.cose.2019.101657).

Pennington, J., Socher, R., and Manning, C. 2014. "Glove: Global Vectors for Word Representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar: Association for Computational Linguistics, pp. 1532–1543. (https://doi.org/10.3115/v1/D14-1162).

Popat, K., Mukherjee, S., Strötgen, J., and Weikum, G. 2016. "Credibility Assessment of Textual Claims on the Web," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management - CIKM '16*, Indianapolis, Indiana, USA: ACM Press, pp. 2173–2178. (https://doi.org/10.1145/2983323.2983661).

Popat, K., Mukherjee, S., Yates, A., and Weikum, G. 2018. "DeClarE: Debunking Fake News and False Claims Using Evidence-Aware Deep Learning," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium: Association for Computational Linguistics, October, pp. 22–32. (https://doi.org/10.18653/v1/D18-1003).

Qazvinian, V., Rosengren, E., Radev, D. R., and Mei, Q. 2011. "Rumor Has It: Identifying Misinformation in Microblogs," in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, Edinburgh, Scotland, UK.: Association for Computational Linguistics, July, pp. 1589–1599. (https://www.aclweb.org/anthology/D11-1147).

Zhou, C., Sun, C., Liu, Z., and Lau, F. C. M. 2015. "A C-LSTM Neural Network for Text Classification," *ArXiv:1511.08630 [Cs]*. (http://arxiv.org/abs/1511.08630).