Deep Learning for Geolocating Social Media Users and Detecting Fake News

Nikos Deligiannis[1,2], Tien Do Huu[1,2], Duc Minh Nguyen[1,2], Xiao Luo[1]

[1]Vrije Universiteit Brussel (VUB), Department of Electronics and Informatics (ETRO), Pleinlaan 2, B-1050 Brussels, BELGIUM

[2]imec, Kapeldreef 75, B-3001 Leuven, BELGIUM

{ndeligia, thdo, mdnguyen}@etrovub.be, xiao.luo@vub.be

ABSTRACT

Social networks like Facebook and Twitter have become a great tool for connecting people, communicating and disseminating ideas. While these platforms are obviously beneficial to our society, concerns are growing with respect to spreading fake news. Such untrustworthy information is likely to create adverse impact on the public opinion, leading to uncertain outcomes of public events such as elections. However, the process of identifying fake news is time-consuming, laborious, and requires expert knowledge. On top of that, geolocating users on social media platforms is a very useful as it enables applications including event detection and social unrest forecasting. Nonetheless, not many users disclose their location via their profile, or the provided information is often unreliable.

As such, news verification and user geo-location on social networks are two important, interconnected big data problems. Machine learning solutions have recently been proposed to address these problems. Most of the existing solutions focus on engineering features extracted from content, social users and information propagation patterns. In this work, we will present novel solutions to fake news detection and user geolocation using deep neural networks. The results from the experiments conducted on several real-world datasets show that our methods outperform the existing approaches.

1.0 INTRODUCTION

Over the last decade, social media platforms like Twitter and Facebook have become a very important tool for connecting people. Being able to provide open and quick access to a sheer amount of information, social networks have attracted hundreds of millions of users across the globe. According to [1], Facebook has more than 2.2B active users by March 2018, while the number of active users on Twitter is more than 330M. The information on social networks is user-generated; therefore, it is greatly diverse. People talk about everything via social posts, and the posts could be shared or republished. Using this mechanism, the information can be propagated widely in a short time. For this reason, social media platforms can be seen as human sensing networks.

Among social media platforms, Twitter seems to be the most attractive for researchers because of its open data application programming interfaces (APIs). On Twitter, users publish short text messages, known as tweets. Tweets can be re-published by other users, which are called retweets. These messages and user profiles can be easily collected via the Twitter API; as such, a lot of popular datasets for social network analysis research are collected from Twitter.

Unverified information like rumors or fake news find social networks a breeding ground. Fake news, intentionally written for financial incentive, politic influence or other purposes is very difficult to detect. It often requires expert knowledge to judge a meticulously manipulated piece of misleading news. Therefore, online users, who usually just skim posts on social networks, are likely to spread fake news. Fake news is believed to be a severe social problem because it causes harm not only to a single person but also to societies and countries. More precisely, fake news might have adverse effect on finance of giant corporations, social inter-class contradictions, propagating hatred and distrust among citizens. As an example, in September

2016, a black man in the Boston area was allegedly to be killed by a police officer named Thomas Wright for merely refusing to put out his marijuana cigarette. The reported unfair treatment of police officers to black men once stimulated the rage of black man communities towards the local police office and the white people in the U.S. However, this incident was judged as a fake news in the following days [2]. Moreover, fake news is believed to play a role in the US president election in 2016 [3]. These examples, and many others, advocate that an efficient method for automatic fake news detection is desperately needed.

A strongly related problem is user geolocation. The location of users on social networks is a very useful piece of information, enabling many applications such as location-based service recommendation [4], social unrest forecasting [5], and migration flow analysis. Concerning fake news, it is helpful to know the location of fake news purveyors and propagators. However, the users' location is not always available. Furthermore, this information might be very confusing and ambiguous. Therefore, plenty of efforts have been spent on developing techniques for user geolocation.

In this paper, we introduce novel methods for automatic fake news detection and user geolocation on Twitter using deep neural networks, known as deep learning[1]. The paper builds on recent results from our research group [40], [41], where we train deep models on publicly available Twitter datasets. The proposed deep models outperform state-of-the-art methods in both the problem of Twitter user geolocation and fake news detection.

The rest of this paper is organized as follows. In Section 2, we briefly summarize related work. In Section 3, we present our methods for fake news detection and user geolocation. In Section 4 we present the experimental results, and then we draw the conclusion in Section 5.

2.0 RELATED WORKS

There are several approaches in detecting fake news. One approach is based on the news content while an alternative one utilizes the social context of the news. The former solely considers the textual content feature and existing factual sources to classify the news; this approach can be further categorized into knowledge-based and style-based methods. Knowledge-based methods look at the external sources to verify the claims in the news content [9, 10]. On the other hand, style-based methods detect the particular writing styles in the misleading content [11, 12]. The social context-based approach employs the relevant engagements of users on social networks. Two sub-categories of this approach include the stance-based and propagation-based approaches. In the stance-based methods, the viewpoints of users are captured for supporting fake news analysis [13, 14]. In propagation-based methods, on the other hand, the pattern in the interrelations of relevant social media posts are considered [15, 16]. Our method incorporates the news content and the social context features using a unique model, namely, a graph convolutional neural network [26].

Similar to the problem of fake news detection, there are two main approaches for geolocating Twitter users. The first approach relies on the textual expressiveness of tweets. This approach has been studied thoroughly in many previous publications [17, 18, 19]. The second approach looks at the online relationship of Twitter users. The basic idea is that people often interact with others in their vicinity; therefore, the location of a user can be estimated via the locations of his/her friends [20, 21, 22]. Metadata can also be exploited to predict users' locations [23]. Alternative works have already combined two approaches [24, 25]; however, these proposed methods are not straightforward, requiring multiple steps of fine-tuning. Namely, these approaches [25] adhere to a two-step approach: (*i*) they build a model to classify disparate users in a geographical area and (*ii*) they apply label propagation to predict the locations for all users. In contrast, our method [40], [41] combines the two main approaches using a unified model following the multi-view deep learning paradigm. As a result, our method is more scalable and easier to train than existing solutions.

[1] Deep learning has been proven very effective in many applications areas such as image classification [6], speech recognition [8], compressed sensing [7], and matrix factorization [42].

3.0 PROPOSED METHODS

3.1 Fake news detection using a graph convolutional neural network

In fake news detection the goal is to classify the event described in the news into fake or real, which is a binary classification problem. We propose to use the graph convolutional neural network (GCN) [26] to address the classification problem. The motivation of using a GCN is that we want to exploit the relation between events and publishers. Concretely, if a publisher spreads a fake news before, it is likely that he will repeat this action. On the other hand, GCNs have been proven to perform very well on relational structured data. In order to leverage the strength of GCNs, we first build a graph of events. The classification of events will be done by the GCN directly on this graph.

3.1.1 Graph establishment

There are three entities involved in the fake news life cycle: the publishers, the events and the users on social networks. Fig. 1 shows the relationship among them. By exploiting this relationship to infer the correlation between the labelled and unlabelled events, we can classify the unlabelled. Some publications have already explored this concept using label propagation-based techniques [15, 27]. In our work, instead of using these label propagation algorithms, we employ a GCN [26] to do the prediction.

We consider an event as a node and we make the edges as follows: If two articles are published by the same publisher, a connection between them is established. Furthermore, if a social user engages two events, these articles should also be connected. The weight of an edge is the number of times two articles are connected. Fig. 1 shows how the news graph is created. We can see that event e_1 is connected to event e_2 because they are published by the same publisher p_1. Events e_3 and e_4 are linked because user u_3 refers to these events.

3.1.2 Graph convolutional network architecture

GCNs enable supervised classification on graph-like data. Fig. 2 shows the general architecture of a graph convolutional neural network (GCN). Similar to classical neural networks, a GCN has an input layer, hidden layers and an output layer. The input layer receives the C-channel signal of the graph, then the signal is transformed at the hidden layers. Finally, at the output layer, the signal is converted to probabilities using the SoftMax function.

For news classification, we build a simple GCN architecture with only one hidden layer. The chosen objective function is the cross-entropy. The weights of the model are trained using the stochastic gradient descent algorithm. The signal on each node is the feature vector extracted from the corresponding event's description. Concretely, we compute *word2vec* embeddings [28] for each term in the news content and calculate the average of these embeddings to obtain the representation for the event. This weighting scheme produces a simple but expressive feature vector for event classification.

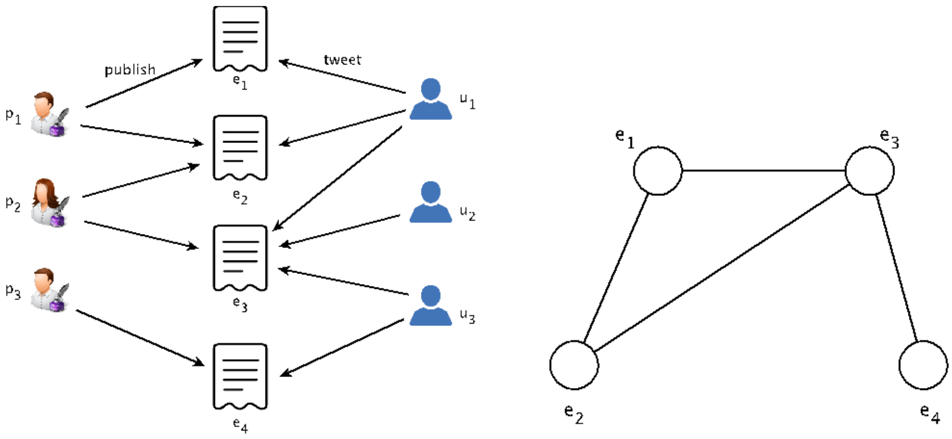

Fig. 1. Three considered entities in news propagation (left) and the graph of events (right).

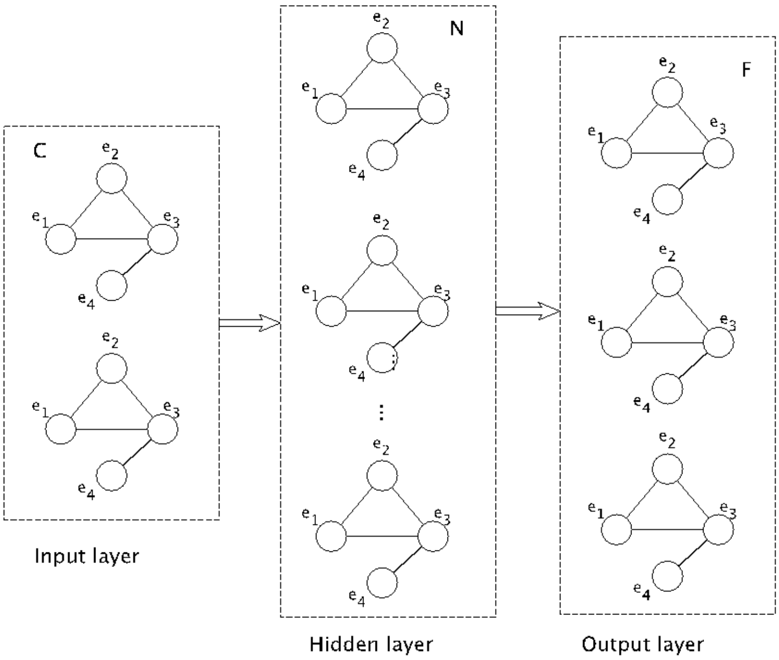

Fig. 2. Schema of a Graph convolutional neural network architecture. C is number of input channels (e.g. a C-dimensional feature vector of each node), F is number of feature maps in the output layer and N is the size of the hidden layer.

3.2 Geolocation of Twitter users using Multiview learning

We consider the task of predicting the continuous geographical coordinates of Twitter users. To achieve this goal, we propose a multi-entry neural network architecture (MENET) that can handle different categories of features [40], [41]. The network can be unbounded at the output layer to directly obtain the coordinates. However, we convert the task to classification as it is the typical use of neural networks. Concretely, we use administrative boundaries as the labels for classification. More precisely, the labels are the states in the US because the considered datasets are collected from Twitter users who reside in the US. When the state of a user is predicted, we assign the centroid's coordinates of the state to the user. The centroid's coordinates are the median of all known coordinates of Twitter users available from the training set.

In order to predict the user's location, tweets from that user are collected. Tweets from the same user are then concatenated making up a tweet document. The textual content of the document might provide clues for the location of the user. Moreover, using the textual content, the online relationship of the user can be found, revealing their location. Metadata can also be exploited for geolocating users. In the next sections, we present our model architecture. After that, we show the features used for realizing the proposed model. For more details on the model, the features, and the encoding at the output layer we refer to [40], [41].

3.2.1 Model architecture

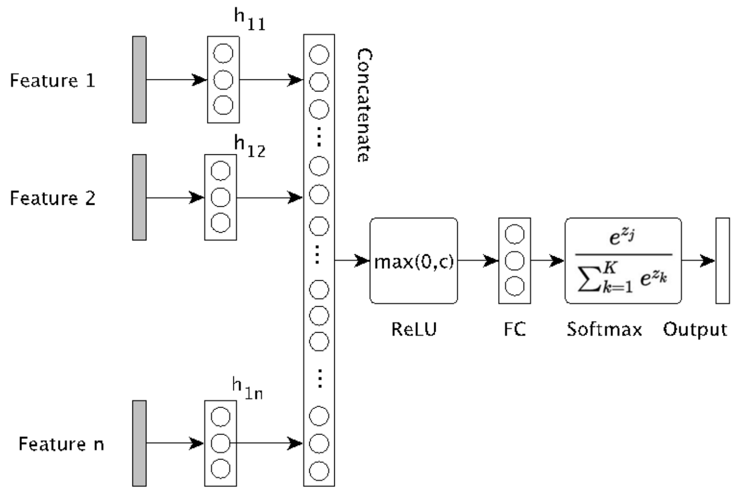

Fig. 3. The architecture of our multi-entry neural network model for Twitter user geolocation.

Our model architecture is presented in Fig. 3. The architecture is able to accept multiple features via its branches corresponding to different views of Twitter data. Each branch has a separate hidden layer. In the learned feature space, the activation outputs from these hidden layers are concatenated. To add non-linearity to the network, we use the ReLU function [29]. The SoftMax function is employed just before the output to obtain class probabilities. Considering *n* classes of users [corresponding to states in the US], we use the cross-entropy loss as the objective function. The network is trained using stochastic gradient descent (SGD). To prevent overfitting, we employ weight decay, dropout and early stopping as regularisation strategies. We use the provided development set to fine-tune the network's parameters. In the testing phase, we measure the performance of MENET using the geolocation accuracy.

3.2.2 Multiview features

Term frequency-inverse document frequency: Also known as TF-IDF, this is a feature at lexical level. TF-IDF is a numerical statistic, intended to reflect how important a term is to a document in a text corpus. It is proportional to the frequency of a term in the document (TF), counteracted by how often the term occurs in the corpus (IDF).

Document embedding: TF-IDF captures the most indicative terms in a document without understanding the context of the whole document. To obtain the vector representation of a document, we employ *doc2vec,* known as Distributed Representation of Sentences [30]. Using tweet documents from the users available in the training set, we train a doc2vec model. The embeddings of the tweet documents representing Twitter users are then inferred from the model. Details on the training and inferring processes are presented in Section 4.2 and can also be found in [40], [41].

Node embedding: The TF-IDF and document embeddings are word and semantics level features; therefore, they do not represent the relationship between users. In order to extract this kind of feature, we build a graph of users, and employ the *node2vec* algorithm [31] to extract node embeddings, which encode the local graph structure. The graph of Twitter users is built based on [24, 25]. In the graph, a node represents a user while an edge shows the interaction between two users. Concretely, if a user mentions another user, we make an edge between them. If two users mention an external user, a connection between the two is created. We make the graph less dense by using a *celebrity* threshold. A node is considered as celebrity when its degree is larger than the threshold. The connections to the celebrity nodes are removed.

Timestamp embedding: The work in [32] shows that there is a relation between the time and the location of a Twitter stream. In order to exploit this characteristic, we also extract timestamp embeddings. A timestamp embedding is a vector of 24 dimensions, and each element is the total number of tweets posted in the corresponding hour. The variation in the posting activity provides a clue for the longitude of the user's location.

4.0 EXPERIMENTAL RESULTS

4.1 Datasets

We employ several popular datasets for our experiments. For user geolocation experiments, we use GeoText [33] and UTGeo2011 [34]. For fake news classification, we conduct experiments on FakeNewsNet [35].

GeoText: This is a small dataset containing more than 300K tweets collected from 9475 Twitter users in the US. All the tweets are geo-tagged and the only metadata available is the posting time of the tweets (timestamp). The dataset is already split into the training, development and test sets.

UTGeo2011: This dataset is much larger than GeoText with approximate 38M tweets harvested from more than 400K Twitter users from the US. Similar to GeoText, all tweets are geo-tagged, and the timestamps are also available. The splitting of this dataset is given.

FakeNewsNet: It consists of two small datasets, namely BuzzFeed and PolitiFact. The names suggest the news items are collected from the *buzzfeed* and *politifact* fact-checking websites. Both dataset contain the content and labels for the news items. For each news item, a number of social engagements, which are tweets, is included. Table 1 shows statistics for these fake news datasets. Following [35], 80% of the data is kept for training and parameter fine-tuning, and the rest is used for testing. We conduct experiments with 5-fold cross validation.

4.2 Implementation details

The implementation consists of three phases: Data pre-processing, feature extraction and model building. The data pre-processing step is necessary because our datasets contain plenty of tweets, which are noisy. Therefore, tweets need to be tokenized and functional words are removed. Another procedure, which is stemming, is applied to obtain the original form of words. These procedures are performed using the NLTK library [36].

As mentioned before, we use four types of features in geolocating Twitter users, namely TF-IDF, *doc2vec*, *node2vec* and timestamp. To obtain the TF-IDF feature, we use the existing *scikit-learn* library [37], whereas, to extract the doc2vec feature, we use the genism library [38]. Finally, node2vec embeddings are constructed using the code from [31]. In fake news detection, we employ the averaged word2vec.

We employ the Tensorflow deep learning environment to implement the proposed models. For the MENET architecture, we use tensorflow to build the model and train the model using Adam algorithm [39]. The code for the graph convolutional network [26] is used to train the fake news classifier.

Table 1. Statistics for BuzzFeed and PolitiFact

	BuzzFeed	PolitiFact
True news	91	120
Fake news	91	120
Number of users	15.257	23.865
Number of engagements	25.240	37.259
Number of publishers	9	91

To evaluate the geolocation accuracy, we employ the mean and median distance error, as well as the accuracy within 161 km (~100 mile) known as @161. The mean and median distance errors are measured in kilometres (km). The distance between the predicted location and the ground-truth location is calculated using the Haversine formula. The @161 metric is expressed in percentage (%). Concerning fake news detection, which is a binary classification problem, the typical performance measurement is the accuracy. Moreover, the considered datasets are balanced. Therefore, it is not necessary to compute other metrics such as precision or recall.

4.3 Result

Tables 2 and 3 show the experimental results for fake news detection and Twitter user geolocation. The fake news classification result shows that the use of the graph convolutional neural networks for fake news detection is very promising. Compared to the previous work of Shu et al. [35], we achieve much better accuracy only with one-hidden-layer GCN architecture. Moreover, the GCN is easier to scale than the optimization method in [35] and training a GCN is relatively fast on a GPU.

Concerning the geolocation problem, we achieve the lowest mean distance error on both GeoText and UTGeo2011 datasets against all competing methods (see Table 3). Moreover, we obtain slightly lower median distance error and higher @161 accuracy on the GeoText dataset. However, the median error and @161 on the UTGeo2011 are not as good as the previous works. This is understandable because we did not consider the distribution of Twitter users. In our recently extended model [41]–a.k.a., the MENET with S2 labels–which uses a state-of-the-art partitioning scheme [43], we significantly improve the location.

5.0 CONCLUSION

Detecting fake news and rumors in social media platforms is a critical issue. Plenty of efforts have been spent on detecting fake news automatically; however, no methods have achieved very high accuracy. Fake news detection is highly related to the problem of user geolocation in social networks. Understanding where the fake news appears will allow tracing their propagation. In this work, we present two methods for detecting fake news and predicting the location of users using deep learning techniques. The results show that our methods are capable to address the considered tasks efficiently, surpassing the performance reported in previous works. We are working towards improving further the accuracy of the proposed models and performing experiments on large datasets.

Table 2. Fake news classification on BuzzFeed and PolitiFact

	BuzzFeed	PolitiFact
Shu et al. [35]	0.864	0.878
GCN	**0.944**	**0.895**

Table 3. Geolocation accuracy on GeoText and UTGeo2011

	GeoText			UTGeo2011		
	Mean (km)	Median (km)	@161 (%)	Mean (km)	Median (km)	@161 (%)
Liu and Inkpen [18]	855.9	N/A	N/A	733	377	24.2
Cha et al. [17]	581	425	N/A	N/A	N/A	N/A
Rahimi et al [24]	581	**57**	59	529	78	60
Rahimi et al [25]	578	61	59	515	**77**	**61**
Our MENET [40]	**570**	58	**59.1**	**474**	157	50.5
Our MENET with S2 labels [41]	**532**	**32**	**62.3**	**433**	**45**	**66.2**

6.0 REFERENCES

[1] Statista. (2017, Nov) Number of monthly active twitter users worldwide. [Online]. Available: https://www.statista.com/statistics/282087/number-of- monthly-active-twitter-users/

[2] Snopes. (2016, Sept) Boston Police Officer Kills Black Man Over Marijuana Cigarette. [Online]. Available: https://www.snopes.com/fact-check/boston-police-officer-kills-black-man-over-marijuana-cigarette/

[3] Business Insider. (2016, Nov) This is what fake news actually looks like. [Online]. Available: http://uk.businessinsider.com/fake-presidential-election-news-viral-facebook-trump-clinton-2016-11?r=UK&IR=T/#11-sarah-palin-banned-muslims-from-entering-her-daughter-1

[4] J. Bao, Y. Zheng, D. Wilkie, and M. Mokbel, "Recommendations in location-based social networks: a survey," *GeoInformatica*, vol. 19, no. 3, pp. 525–565, 2015.

[5] R. Compton, C. Lee, T. Lu, L. D. Silva, and M. Macy, "Detecting future social unrest in unprocessed twitter data: emerging phenomena and big data," in *IEEE International Conference on Intelligence and Security Informatics*, 2013, pp. 56–60.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[7] D. M. Nguyen, E. Tsiligianni, N. Deligiannis, "Deep learning sparse ternary projections for compressed sensing of images", in IEEE Global Conference on Signal and Information Processing, GlobalSIP'17, Montreal, Canada, November 2017.

[8] A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 6645–6649.

[9] Amr Magdy and Nayer Wanas. Web-based statistical fact checking of textual documents. In Proceedings of the 2nd international workshop on Search and mining user-generated contents, pages 103–110. ACM, 2010.

[10] You Wu, Pankaj K Agarwal, Chengkai Li, Jun Yang, and Cong Yu. Toward computational fact-checking. Proceedings of the VLDB Endowment, 7(7):589–600, 2014

[11] Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. A stylometric inquiry into hyper-partisan and fake news. arXiv preprint arXiv:1702.05638, 2017.

[12] Victoria L Rubin and Tatiana Lukoianova. Truth and deception at the rhetorical structure level. Journal of the Association for Information Science and Technology, 66(5):905–917, 2015.

[13] Eugenio Tacchini, Gabriele Ballarin, Marco L Della Vedova, Stefano Moret, and Luca de Alfaro. Some like it hoax: Automated fake news detection in social networks. arXiv preprint arXiv:1704.07506, 2017

[14] Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. News verification by exploiting conflicting social viewpoints in microblogs. In AAAI'16

[15] Zhiwei Jin, Juan Cao, Yu-Gang Jiang, and Yongdong Zhang. News credibility evaluation on microblog

with a hierarchical propagation model. In ICDM'14.

[16] Manish Gupta, Peixiang Zhao, and Jiawei Han. Evaluating event credibility on twitter. In PSDM'12

[17] M. Cha, Y. Gwon, and H. T. Kung, "Twitter geolocation and regional classification via sparse coding," in *International AAAI Conference on Web and Social Media*, 2015, pp. 582–585.

[18] J. Liu and D. Inkpen, "Estimating user location in social media with stacked denoising auto-encoders." in *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2015, pp. 201–210.

[19] R. Priedhorsky, A. Culotta, and S. Y. D. Valle, "Inferring the origin locations of tweets with quantitative confidence," in *ACM conference on Computer supported Cooperative Work & Social Computing*, 2014, pp. 1523–1536.

[20] L. Backstrom, E. Sun, and C.Marlow, "Find me if you can: improving geographical prediction with social and spatial proximity," in International Conference on World Wide Web, 2010, pp. 61–70.

[21] D. Jurgens, "That's what friends are for: Inferring location in online social media platforms based on social relationships." in International AAAI Conference on Weblogs and Social Media, 2013, pp. 273–282.

[22] R. Compton, D. Jurgens, and D. Allen, "Geotagging one hundred million twitter accounts with total variation minimization," in IEEE International Conference on Big Data, 2014, pp. 393–401.

[23] J. H. Lau, L. Chi, K. N. Tran, and T. Cohn, "End-to-end network for twitter geolocation prediction and hashing," in *International Joint Conference on Natural Language Processing*, 2017.

[24] A. Rahimi, T. Cohn, and T. Baldwin, "A neural model for user geolocation and lexical dialectology," in *Meeting of the Association for Computational Linguistics*, 2017, pp. 209–216.

[25] A. Rahimi, T. Cohn, and T. Baldwin, "Twitter user geolocation using a unified text and network prediction model," in *Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing*, 2015, pp. 630–636.

[26] T. N. Kipf, and M. Welling. "Semi-supervised classification with graph convolutional networks." *arXiv preprint arXiv:1609.02907* (2016).

[27] K. Wu, S. Yang, and K. Q. Zhu. "False rumors detection on sina weibo by propagation structures". In *2015 IEEE 31st International Conference on Data Engineering (ICDE)*, 2015, pp. 651-662.

[28] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. "Distributed representations of words and phrases and their compositionality." In *Advances in neural information processing systems*, pp. 3111-3119. 2013.

[29] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *International Conference on Artificial Intelligence and Statistics*, 2011, pp. 315–323.

[30] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *International Conference on Ma- chine Learning*, 2014, pp. 1188–1196.

[31] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *ACM SIGKDD*

International Conference on Knowledge Discovery and Data Mining, 2016, pp. 855– 864

[32] M. Dredze, M. Osborne, and P. Kambadur, "Geolocation for twitter: Timing matters.," in Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016, pp. 1064– 1069.

[33] J. Eisenstein, B. O'Connor, N. A. Smith, and E. P. Xing, "A latent variable model for geographic lexical variation," in *Conference on Empirical Methods in Natural Language Processing*, 2010, pp. 1277–1287.

[34] S. Roller, M. Speriosu, S. Rallapalli, B. Wing, and J. Baldridge, "Supervised text-based geolocation using language models on an adaptive grid," in *Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 2012, pp. 1500–1510.

[35] K. Shu, S. Wang, and H. Liu. "Exploiting Tri-Relationship for Fake News Detection." *arXiv preprint arXiv:1712.07709* (2017).

[36] S. Bird, E. Loper and E. Klein (2009), *Natural Language Processing with Python*. O'Reilly Media Inc.

[37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Ma- chine learning in Python," *Journal of Machine Learning Re- search*, vol. 12, pp. 2825–2830, 2011.

[38] R. Rehurek and P. Sojka, "Software framework for topic modelling with large corpora," in LREC 2010 Workshop on New Challenges for NLP Frameworks, 2010.

[39] D.P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *The International Conference on Learning Representations*, 2015.

[40] T. Do Huu, D. M. Nguyen, E. Tsiligianni, B. Cornelis, N. Deligiannis, "Twitter user geolocation using deep multiview learning", in IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'18, Calgary, Alberta, Canada, April 2018.

[41] T. Do Huu, D. M. Nguyen, E. Tsiligianni, B. Cornelis, N. Deligiannis, "Multiview deep learning for predicting twitter users' location," arXiv preprint arXiv:1712.08091, 2017.

[42] D. M. Nguyen, E. Tsiligianni, N. Deligiannis, "Learning discrete matrix factorization models", IEEE Signal Processing Letters, 2018.

[43] S2Geometry, GitHub repository, https://github.com/google/s2geometry