Sentiment based Graph Learning for Fake News Detection and Analysis

Saptarashmi Bandyopadhyay
sbandyo20@gmail.com
The Pennsylvania State University
University Park, Pennsylvania, U.S.A.

Laxmaan Balaji
lpb5347@psu.edu
The Pennsylvania State University
University Park, Pennsylvania, U.S.A.

Manasi Dinesh Nair
mnsdnair@gmail.com
The Pennsylvania State University
University Park, Pennsylvania, U.S.A.

ABSTRACT

Fake news originating from social media, news websites, blogs among other information sources affects our daily discourse by creating a misleading narrative. It has been reported by [9] that 67% citizens of U.S.A. are primarily receiving their news from social media in 2017 due to fast communication, easy access and low cost. Satirical news articles with humor content make it all the more difficult to identify fake news solely based on the intention of the creator and the perception of the reader. Current research on automated detection of fake news has primarily focused on validation using background knowledge base, identifying the writing style of articles sharing false information based on linguistic features, analyzing the propagation of deceptive information, and evaluating the credibility of ambiguous content. There are several challenges to this approach namely early detection of fake news, identification of false news across multiple domains, and weakly-supervised and semi-supervised fake news detection. Our work has enhanced existing research on automated detection and analysis of fake news across 6 domains namely business, education, entertainment, politics, sports and technology by graphical model-based sentiment analysis of the news articles.

A graph dataset of news articles from 6 domains in [8] has been created with each news article, either legitimate or fake as a node, composite sentiment scores of each article as node features and edge connectivity between news articles across domains. This edge connectivity approach, crucial to model the inter-dependent sentiments in legitimate and fake news across different domains, has led to a high classification accuracy of legitimate and fake news at 93.75% over the existing SVM based state-of-the-art score of 74% in the *FakeNewsAMT Dataset*. Cross-domain analysis of fake news detection has been efficiently undertaken by this approach with the calculation of precision and F-measures. Experiments have been carried out to demonstrate that the Graph Neural Network based approach is far more efficient than applying SVM or Random Forest classification with the same set of sentiment features.

IST 597 Deep Learning Class, Pennsylvania State University, University Park, August–December, 2019, University Park, PA

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

CCS CONCEPTS

• **Computing methodologies → Neural networks**; *Classification and regression trees*; *Support vector machines*; • **Computer systems organization → Neural networks**; • **Information systems →** *Retrieval tasks and goals.*

KEYWORDS

graph neural networks, fake news detection, sentiments, edge connectivity, graph dataset creation

ACM Reference Format:
Saptarashmi Bandyopadhyay, Laxmaan Balaji, and Manasi Dinesh Nair. 2018. Sentiment based Graph Learning for Fake News Detection and Analysis. In *IST 597 Deep Learning Class, Pennsylvania State University, University Park, August–December, 2019, University Park, PA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Fake news is now viewed as one of the greatest threats to modern society. Since as many as 62% of U.S. adults consume news on social media (Jeffrey and Elisa, 2016), being able to identify fake content in online sources is a pressing need.

A comprehensive review of detecting fake news on social media, including fake news characterizations on psychology and social theories, existing algorithms from a data mining perspective, evaluation metrics and representative datasets has been presented in [13]. [16] systematically presents the fake news detection strategies from four perspectives (i.e., knowledge, style, propagation, and credibility) and the ways that each perspective utilizes techniques developed in data/graph mining, machine learning, natural language processing, and information retrieval. It suggests that fake news detection from a knowledge perspective is a "comparison" between the relational textual knowledge extracted from to-be-verified news articles and that of knowledge graphs representing facts or ground truth. [12] reviews network properties for studying fake news, introduce popular network types and propose how these networks can be used to detect and mitigate fake news on social media. During fake news dissemination, different entities are involved that can be categorized into content, social and temporal dimensions. In addition, the paper maintains that the inherent network properties motivate and strengthen the need to perform network analysis to study fake news. The dimensions of fake news dissemination revealed mutual relations and dependencies that can form different types of networks. [10] suggests, there are three generally agreed upon characteristics of fake news: the text of an article, the user response it receives, and the source users promoting it. They propose a model that combines all three characteristics for a more accurate and automated prediction. Experimental analysis of real-world data demonstrated that their model extracts meaningful latent representations of both users and articles.

[1] describes random forests as a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. . The generalization error for forests converges a.s. to a limit as the number of trees in the forest becomes large. It argues that injecting the right kind of randomness makes them accurate classifiers and regressors.

[5] present VADER, a simple rule-based model for general sentiment analysis, and compare its effectiveness to eleven typical state-of-practice benchmarks including LIWC, ANEW, the General Inquirer, SentiWordNet, and machine learning oriented techniques relying on Naive Bayes, Maximum Entropy, and Support Vector Machine (SVM) algorithms. They found that VADER outperforms individual human raters (F1 Classification Accuracy = 0.96 and 0.84, respectively), and generalizes more favorably across contexts than many benchmarks.

[11] propose a new neural network model, called graph neural network (GNN) model, that extends existing neural network methods for processing the data represented in the graph domain. They furnish a learning algorithm to estimate the parameters of the neural networks based on the backpropagation techniques. This demonstrates its generalization capability. [7] propose an approach of using geometric deep learning on graphs. Their matrix completion architecture combined a novel multi-graph convolutional neural network that can learn meaningful statistical graph-structured patterns from users and items, and a recurrent neural network that applies a learnable diffusion on the score matrix. They showed that this approach compared to traditional methods has low computational complexity and constant number of degrees of freedom independent of the matrix size. They also showed that the use of deep learning for matrix completion beat related state-of-the-art recommender system methods. [15] presents graph attention networks (GATs), novel neural network architectures that operate on graph-structured data, leveraging masked self-attentional layers to address the shortcomings of prior methods based on graph convolutions or their approximations. By stacking layers in which nodes are able to attend over their neighborhoods' features, they enable (implicitly) specifying different weights to different nodes in a neighborhood, without requiring any kind of costly matrix operation (such as inversion) or depending on knowing the graph structure upfront. These models leveraging attention have successfully achieved or matched state-of-the-art performance across four well-established node classification benchmarks, both transductive and inductive (especially, with completely unseen graphs used for testing). [3] developed a hierarchical variational model that introduces additional latent random variables to jointly model the hidden states of a graph recurrent neural network (GRNN) to capture both topology and node attribute changes in dynamic graphs. They argue that adding high level latent variables to graph recurrent neural networks not only increases its expressiveness to better model the complex dynamics of graphs, but also generates interpretable random latent representation for nodes. [14] present an approach to learn representations of graphs using recurrent neural network autoencoders. They trained using sequences generated by random walks, shortest paths, and breadth-first search and demonstrate that their graph representations can increase the accuracy of graph classification tasks on both labeled and unlabeled graphs. [6] proposes an approach for combining the power of high-level spatio-temporal graphs and sequence learning success of Recurrent Neural Networks (RNNs). They made use of factor graph, and factor sharing in order to obtain an RNN mixture that is scalable and applicable to any problem expressed over st-graphs. Their RNN mixture captures the rich interactions in the underlying st-graph. They also demonstrated significant improvements with S-RNN on three diverse spatiotemporal problems including: (i) human motion modeling; (ii) human-object interaction; and (iii) driver maneuver anticipation.

On an operational level, linguistic and fact-checking based approaches have been proposed to discriminate between legitimate and fake news content. Linguistic behaviors such as punctuation usage, word type choices, part-of-speech tags, and emotional valence of a text are involuntary and hard to fake. Hence it can be used as a fake news detection tool. But sophisticated fake news might escape this method. The fact checking approach performs automated verification of propositions to asses truthfulness. But this requires external verification. Therefore, the fact-checking approach is predominantly useful for the detection of deception in texts for which external, verifiable information is available.

To address this, this project uses two novel datasets covering seven different domains are used. The first is collected by combining manual and crowdsourced annotation approaches, the second is collected directly from the web. The crowd sourcing method covered 6 domains namely, sports, business, entertainment, politics, technology and education. The legitimate news comes from mainstream media such as CNN, NYTimes etc and this dataset collected 40 news in each of the six domains, for a total of 240 legitimate news. To ensure the veracity of the news, the dataset also conducted manual fact-checking on the news content, which included verifying the news source and cross-referencing information among several sources. Fake version of the 240 legitimate news comes from crowdsourcing via Amazon Mechanical Turk and manual filters. The web dataset on the other hand, collected news about public figures as they are frequently targeted by rumors, hoaxes, and fake reports. These were collected from online magazines and verified via gossip-checking websites. The dataset collected a total of 500 news articles, with an even distribution for fake and legitimate news. The data was collected in pairs, with one article being legitimate and the other fake.

Previous literature [8] uses a linear SVM classifier that is evaluated using five-fold cross-validation, with accuracy, precision, recall, and F-score as performance metrics. Experiments were conducted with different combinations of feature sets to explore their predictive separately and jointly. These feature sets were ngrams, punctuation, psycholinguistic features, readability and grammar. The N-grams are obtained from Bag of Words for every news article. There are 12 types of punctuations like periods, commas, etc which have been used to differentiate deceptive from correct texts. The psycholinguistic features are categorized under 3 feature sets with the LIWC tool. The readability features and syntax refer to content features, readability metrics and Context-Free Grammar based production rules respectively. The results in the literature show that when using all the features on the two datasets, the SVM approach achieves the best accuracies, with 0.74 and 0.76 for the FakeNewsAMT and the Celebrity datasets respectively. For the FakeNewsAMT dataset, the best performing classifiers are the ones that rely on stylistic features(punctuation and readability). Whereas, the classifiers build with the Celebrity dataset show the best performance when using the psycholinguistic features.

Other approaches to fake news detection were introduced by F. Monti, et al.,2019 and A. Hanselowski, et al.,2017.[4] Federico Monti introduces Fake News Detection on Social Media using Geometric Deep Learning. Since fake and legitimate news spread differently on social media, forming propagation patterns was harnessed for the automatic fake news detection. The Propagation-based approaches have multiple advantages compared to their content-based counterparts, among which is language independence and better resilience to adversarial attacks. This allowed a highly accurate (92.7% ROC AUC) fake news detection. A. Hanselowski performs the experiment on a Fake News Challenge Dataset to get a macro averaged F1 score of 0.609.

One approach for embedding nodes as feature vectors is node2vec [2]. In this paper, the authors describe a strategy to represent each node u in a graph $G(V, E)$, as a vector $f(u) \in \mathbb{R}^d$. To encode a node's surrounding topology into its vector representation, the authors employ a modified random walk algorithm that interpolates behaviors of Breadth-First (BFS) and Depth-First (DFS) search algorithms. The authors modify the classic random walk approach by adding hyperparameters that govern the nature of transition between nodes. The first parameter 'return parameter – p ' controls the extent to which the algorithm backtracks to nodes it has visited. The second parameter, 'in-out parameter – q ' controls the exploration behavior. If $q > 1$ the walk explores its immediate neighborhood (BFS like), if $q < 1$ the walk favors outward exploration of nodes not directly connected to the nodes already visited (DFS like). These walks are used to learn the feature vectors for each node. Stochastic Gradient Descent is used to learn the transition function f that maps a node u to its representation $f(u)$. Node2vec additionally supports learning a feature vector for every pair of nodes which is useful in many real world use cases such as predicting connectivity. This mapping $g(f(u), f(v)) \in \mathbb{R}^{d'}$ encodes the features of pairs of nodes. Some choices of g (such as the mean, L1, L2 norms), offer a mapping such that $d' = d$.

2 THE DATA

The Data used is from the FakeNewsAMT Dataset FakeNewsAMT Dataset introduced in [8] with specifications as follows:

- Split into 6 domains :- Sports, Business, Politics, Education, Entertainment and Technology
- Each domain consists of 80 articles with an even split between legitimate and fake articles. (40 legitimate, 40 fake).
- Legitimate news comes from mainstream media; e.g., CNN, NYTimes. Dataset collected 40 news in each of the six domains, for a total of 240 legitimate news. To ensure the veracity of the news, dataset conducted manual fact-checking on the news content, which included verifying the news source and cross-referencing information among several sources
- Fake version of legitimate news comes from crowdsourcing via Amazon Mechanical Turk and manual filters, there are 240 of these articles.

Sample legitimate article and its corresponding fake article are presented below:

```
Sample fake business article

THE BIG DATA CONSPIRACY

Government and Silicon Valley are looking to enslave us. Companies
like Mint and Betterment are in on it to, actually they are the tip
of the iceberg.
Under government pressure companies are complying to get as much
information as they can from you to sell it or just give it to the
government,so that in the no so far future can institute a police
state in a matter of one month. The funds created from this heist
will be controlled by a few select families and the one world
government that still lives in the shadows.
```

```
Sample legitimate business article

Banks and Tech Firms Battle Over Something Akin to Gold: Your Data

The big banks and Silicon Valley are waging an escalating battle over
your personal financial data: your dinner bill last night  your
monthly mortgage payment  the interest rates you pay. Technology
companies like Mint and Betterment have been eager to slurp up this
data  mainly by building services that let people link all their
various bank-account and credit-card information. The selling point
is to make budgeting and bookkeeping easier. But the data is also
being used to offer new kinds of loans and investment products. Now
banks have decided they aren't letting the data go without a fight.
In recent weeks  several large banks have been pushing to restrict
the sharing of this kind of data with technology companies  according
to the tech firms. In some cases  they are refusing to pass along
information  like the fees and interest rates they charge. Both sides
see big money to be made from the reams of highly personal
information created by financial transactions.
```

2.1 Preprocessing

NLTK's Punkt Tokenizer is used to split each article into sentences and further tokenize those sentences. The VADER [5] tool is used to calculate the sentiments of each sentence. VADER provides a vector of sentiments of the form [positive, negative, neutral, compound]. The first three components denote the percentage of the sentence that has positive, negative and neutral sentiments associated with it. The fourth parameter, the *compound* sentiment denotes the sentiment associated with the sentence. The sentiment for the article is calculated as the mean of the sentiments of all the sentences contained within it.

Article #	Pos	Neg	Neutral	Compound	Label
148	0.047	0.303	0.650	-0.811	fake
167	0.025	0.135	0.839	-0.224	fake
248	0.034	0.142	0.823	-0.525	real
361	0.204	0.031	0.765	0.699	real

Table 1: Processed data

The dataset used for our experiments contains 480 rows, 1 per article with the features being the sentiments obtained from VADER. Two sample rows from the dataset are shown in table 1.

3 LEARNING FRAMEWORKS

Our experiments contain 3 main approaches:

3.1 SVM

In order to establish a baseline performance for the Graph Neural Network (introduced in section 3.3) and also to compare approaches with Perez et. al. [8], we train an SVM classifier on the data as presented in section 2.

The dataset is shuffled and 5 fold cross-validation is used. A linear SVM model from the Scikit-Learn package is used as the classifier.

3.2 Random Forest

To assess the performance of ensemble methods on the dataset, we use Scikit-Learn's Random Forest Classifier implementation. Similar to the SVM approach, we shuffle the data and use 5 fold cross-validation.

3.3 Graph Neural Network

To feed the data into a graph neural network (GNN) we follow the following steps:

- Represent the dataset as a graph, defining nodes and the edge lists.
- Generate node embeddings for the nodes, representing each node in the form of a vector.
- using the embeddings as input, perform classification on the data.

4 GRAPH CREATION AND REPRESENTATION

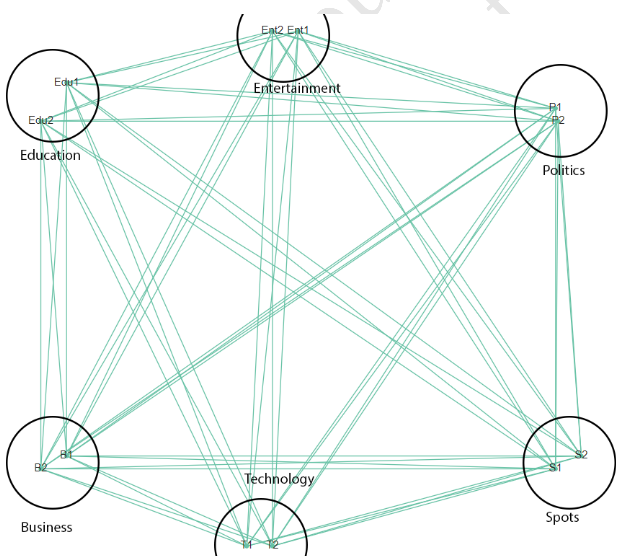

Figure 1: Representation of Nodes in the Graph

The data in section 2 is treated as a list of nodes where the index of the node is given by the article number and the features of the node are the sentiments associated. The node also is labelled as *'real'* or *'fake'* where *'real'* denotes legitimate news article and *'fake'* denotes fake news article.

We come up with a structure of the graph where every node in a particular domain (e.g. Business) is connected to all nodes in the other 5 domains. There are no intra domain connections.

This leads to **480** nodes organized in domains, with no intra domain connections. The total edges in the graph are **59200**.

Fig 1 contains a representation of what the graph would look like if there were only two nodes per domain. The 6 domains have been labeled clearly in the figure. This is just a representation figure and the actual graph is more vast with 78 more nodes in each domain.

A sample excerpt from the edge list is included below:

Sample edge list

```
1       41
1       42
1       43
1       44
1       45
1       46
1       47
...
252       360
252       361
252       362
252       363
252       364
252       365
252       366
...
440       474
440       475
440       476
440       477
440       478
440       479
440       480
```

5 GRAPH NN APPROACH

As outlined in section 3.3, we elaborate on the GNN approach in this section.

5.1 Generating Node Embeddings

Structural properties of the graph like random walks or neighborhood aggregation can be used for node embedding generation In order to generate Node Embeddings, we use the nearest neighbors aggregation method. The sklearn package is used for nearest neighbors implementation. The top **5** neighbors for each node are considered. The dimension of the generated embeddings is **100**. Early stopping is implemented to prevent overfitting. If the validation loss is the same for 10 epochs, then early stipping is enforced.

5.2 The Classifier

The node embeddings are used in the RNN based classification. The classifier is an LSTM based RNN with the following parameters

- 2 layer unidirectional RNN
- 100 hidden units

- Initial Learning rate of 0.001
- Learning rate decay of 0.0001
- L1 regularization
- 0.2 drop-out rate
- Adam Optimizer
- 30 epochs

6 RESULTS

The accuracy, precision, and F measures metrics have been reported in this section on test data. 5-folds cross-validation has been implemented and cross-domain analysis has also been conducted. The confusion matrix has been computed and the training and testing loss plots have also been presented.

6.1 SVM

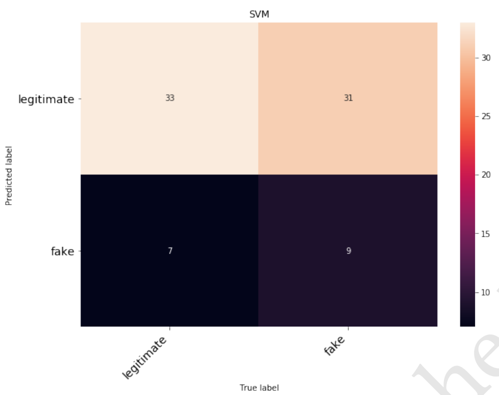

Figure 2: Confusion Matrix for SVM

Our SVM approach using the dataset consisting of 4 features achieved a classification accuracy of **52.5%**.

While our SVM performs worse than the SVM trained by Perez et. al. [8], it is worthwhile to point out that the types and number of features used for both approaches are vastly different. While we used only the sentiments as features, Perez et. al. use a multitude of features such as :

- N-grams
- Punctuations
- Psycho Linguistic Features
- Readability
- Syntax

This surfeit of features contributes to their SVM performing superior to our SVM implementation.

6.2 Random Forest

The random forest approach achieved a classification accuracy of **58.75%**. As pointed out in the previous section, the dearth of features (only 4) is a significant handicap to the Random Forest classifier, giving it a maximum accuracy of 58.75%.

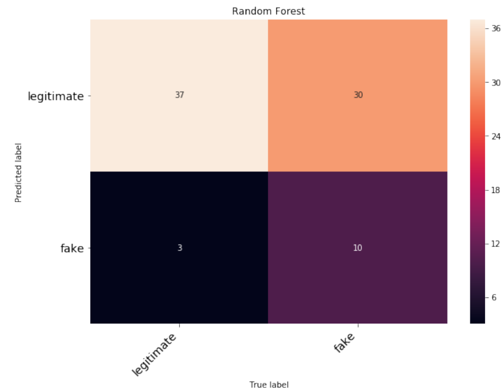

Figure 3: Confusion Matrix for Random Forest

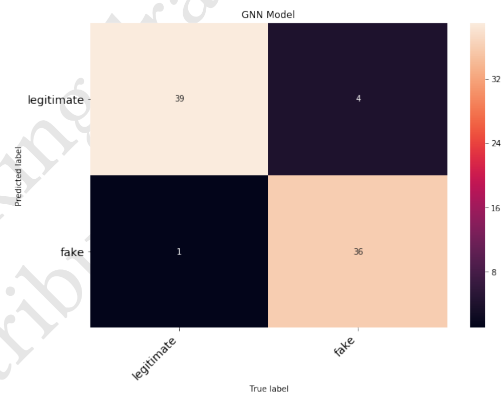

Figure 4: Confusion Matrix for GNN

6.3 GNN

The GNN with the graph dataset based on our node connectivity methodology achieved a classification accuracy of **93.75%**.

6.4 Comparison of Approaches

As can be seen from the confusion matrices, the GNN is the method that is best at detecting fake articles as fake. Both SVM and Random Forest methods suffer from the misclassification of fake articles, leading to low precision. The metrics are summarised in table 2.

Method	Accuracy (%)	Precision (%)	F1 Score (%)
SVM	52.5	51.56	63.46
Random Forest	58.75	55.22	69.16
GNN	**93.75**	**90.7**	**93.96**
SVM as in [8]	74	-	74

Table 2: Comparison of accuracy, precision and F1 score with (Pérez-Rosas et al., 2018)[8] where provided.

To understand the importance of edge connectivity, node embeddings were also generated for a complete graph (all nodes are connected to one another). The accuracy drastically reduced to 67%

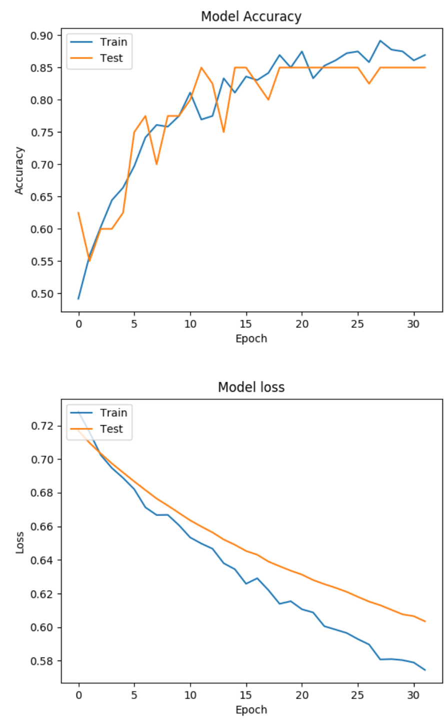

Figure 5: Accuracy and Loss over time for the GNN

then. Thus it can be understood that the correct way to capture the inter-dependence of nodes in the graph through an appropriate edge connectivity scheme is essential for the graph neural network based model to give a successful result.

6.5 Inter-Domain Testing

With the GNN performing well, inter domain testing was done as an experiment where the GNN was trained using 5 fold cross-validation on data from 5 out of 6 domains. The 6th domain with 80 examples (40 legitimate, 40 fake) was used as test data.

As can be seen from the confusion matrices in Fig 6 and 7, the GNN can classify fake news articles from a previously unseen domain with significant accuracy. It beats the performance of the current state of the art methods as shown in table 3.

Amongst all the domains, when held out as test data, we see that the Sports domain had relatively lower performance metrics. The accuracy is **88.75%** and precision **82.98%**. Based on this, one can conclude that **the 'ripples' from other domains do not interact with the Sports domain as much.**

Similarly with the high accuracy **98.75%** and precision **97.56%** in the Education domain, we can infer that the GNN is able to leverage knowledge about other domains and almost perfectly label the test data in the Education domain. Therefore, we can conclude that **news in other domains have strong ties and correlations with the Education domain.**

Figure 6: Confusion Matrices for Inter Domain Testing. Top : Business, Middle: Sports, Bottom: Technology

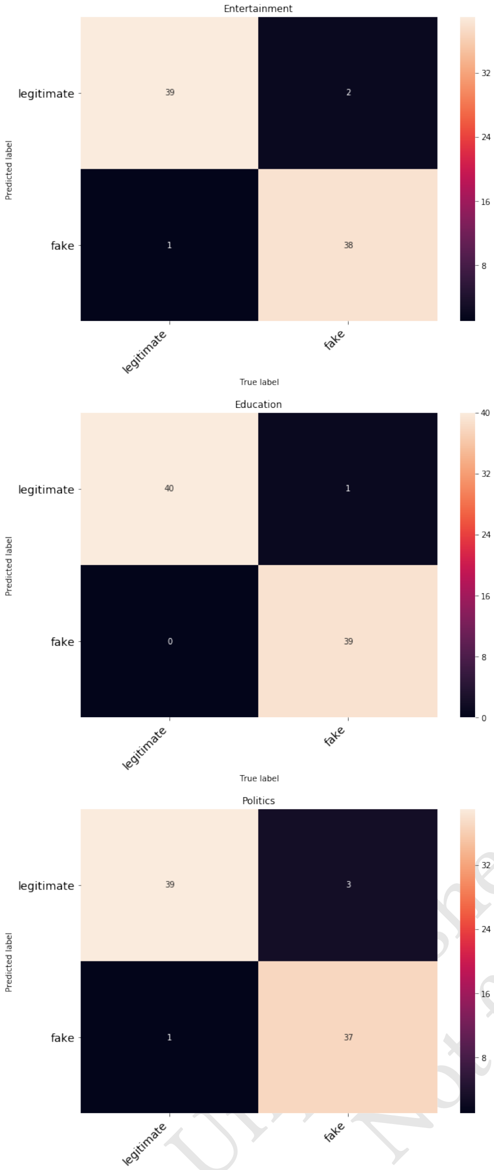

Figure 7: Confusion Matrices for Inter Domain Testing Top : Entertainment, Middle: Education, Bottom: Politics

Domain	Sys Acc	SOTA Acc	Sys Prec	Sys F1 Score
Business	97.5	85	95.23	97.56
Sports	88.75	81	82.98	89.66
Technology	91.25	75	88.37	91.57
Education	98.75	84	97.56	98.77
Entertainment	96.25	75	95.12	96.3
Politics	95	75	92.86	95.12

Table 3: Accuracy, Precision, F1 Score of system vs SOTA Accuracy on each domain

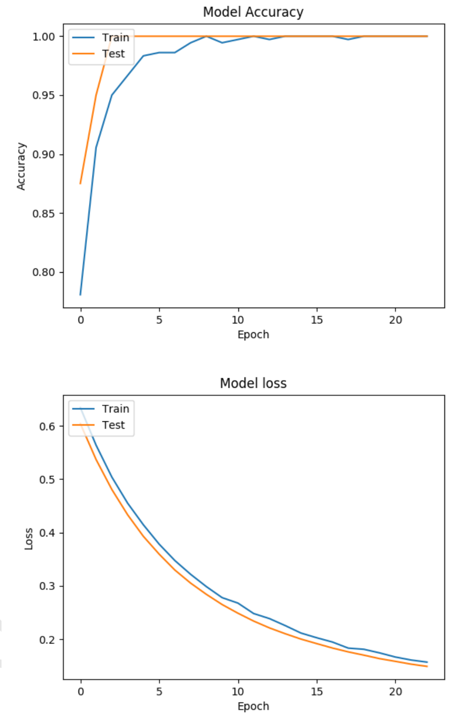

Figure 8: Loss and Acc plots. Test Domain: Business

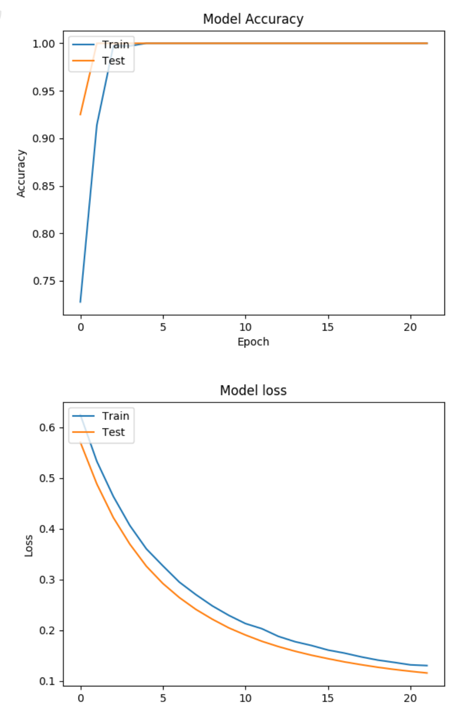

Figure 9: Loss and Acc plots. Test Domain: Education

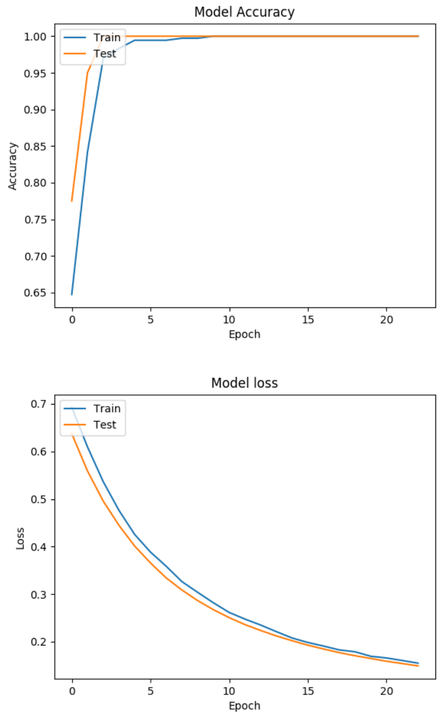

Figure 10: Loss and Acc. plots. Test Domain: Entertainment

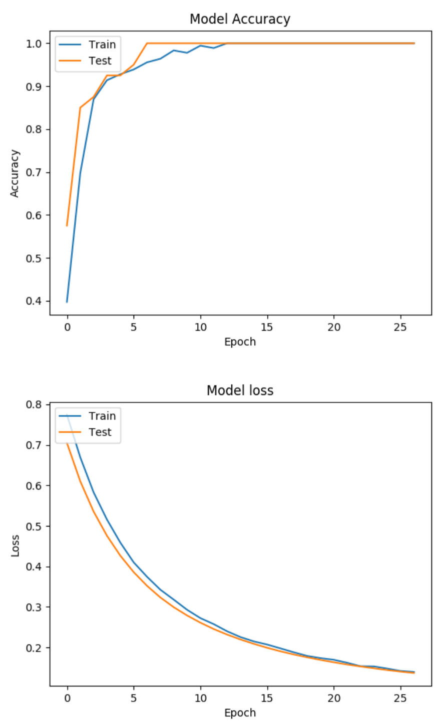

Figure 11: Loss and Acc plots. Test Domain: Politics

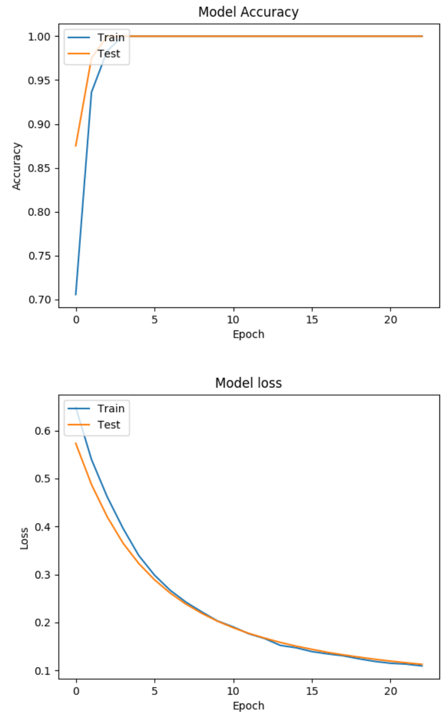

Figure 12: Loss and Acc plots. Test Domain: Sports

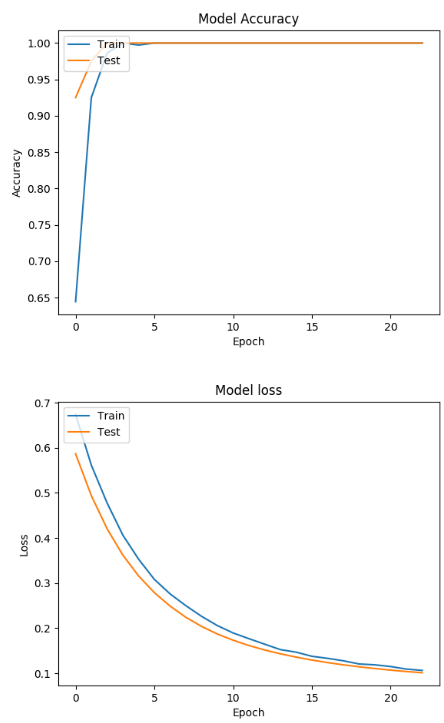

Figure 13: Loss and Acc plots. Test Domain: Technology

7 CONCLUSION AND FUTURE WORK

The sentiment-based graph creation approach has successfully captured the cross-domain dependence of news articles. The classification accuracy with GNN and sentiments is 93.75% which is higher than the state-of-the-art accuracy of 74%. The edge connectivity is very important as a complete graph decreases the classification accuracy. The classification accuracy with GNN is significantly higher with news articles from each of the 6 domains as test datasets than the corresponding state-of-the-art accuracies. This work can be extended to implement knowledge aware graph learning by extracting time period and location as features from the news articles. This approach can address past news articles being treated to have happened in the current time and can improve the accuracy of fake news detection.

REFERENCES

[1] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.

[2] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 855–864.

[3] Ehsan Hajiramezanali, Arman Hasanzadeh, Krishna Narayanan, Nick Duffield, Mingyuan Zhou, and Xiaoning Qian. 2019. Variational graph recurrent neural networks. In *Advances in Neural Information Processing Systems*. 10700–10710.

[4] Andreas Hanselowski, Avinesh PVS, Benjamin Schiller, Felix Caspelherr, Debanjan Chaudhuri, Christian M Meyer, and Iryna Gurevych. 2018. A retrospective analysis of the fake news challenge stance detection task. *arXiv preprint arXiv:1806.05180* (2018).

[5] Clayton J Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.

[6] Ashesh Jain, Amir R Zamir, Silvio Savarese, and Ashutosh Saxena. 2016. Structural-RNN: Deep learning on spatio-temporal graphs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5308–5317.

[7] Federico Monti, Michael Bronstein, and Xavier Bresson. 2017. Geometric matrix completion with recurrent multi-graph neural networks. In *Advances in Neural Information Processing Systems*. 3697–3707.

[8] Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2017. Automatic detection of fake news. *arXiv preprint arXiv:1708.07104* (2017).

[9] KaiShu HuanLiu RezaZafarani, XinyiZhou. 2019. Fake News Research: Theories, Detection Strategies, and Open Problems. https://www.fake-news-tutorial.com/. Accessed: 2019-09-01.

[10] Natali Ruchansky, Sungyong Seo, and Yan Liu. 2017. Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 797–806.

[11] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. 2008. The graph neural network model. *IEEE Transactions on Neural Networks* 20, 1 (2008), 61–80.

[12] Kai Shu, H Russell Bernard, and Huan Liu. 2019. Studying fake news via network analysis: detection and mitigation. In *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*. Springer, 43–65.

[13] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter* 19, 1 (2017), 22–36.

[14] Aynaz Taheri, Kevin Gimpel, and Tanya Berger-Wolf. 2018. Learning graph representations with recurrent neural network autoencoders. *KDD Deep Learning Day* (2018).

[15] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).

[16] Reza Zafarani, Xinyi Zhou, Kai Shu, and Huan Liu. 2019. Fake News Research: Theories, Detection Strategies, and Open Problems. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM.