Semi-supervised Content-based Fake News Detection using Tensor Embeddings and Label Propagation

Gisel Bastidas Guacho
UC Riverside
gbast001@ucr.edu

Sara Abdali
UC Riverside
sabda005@ucr.edu

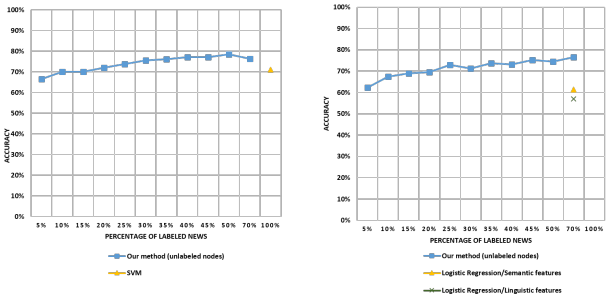
Evangelos E. Papalexakis
UC Riverside
epapalex@cs.ucr.edu

Abstract

Fake news detection is an emerging problem that has become extremely prevalent during the last year. In this paper, we focus on content-based detection of fake news articles, while assuming that we have a small amount of labels. To that end, we leverage tensor decomposition to derive concise article embeddings that capture spatial/contextual information about each article, and use those embeddings to create a graphical representation of the articles, on which we propagate the limited amount of labels available to us. Results on three real-world datasets show that our method perform on par or better than existing models that are fully supervised.

1 Introduction

Misinformation on the web is a problem that has been greatly amplified by the use of social media, and the problem of fake news, in particular, has become ever more prevalent during the last year. Fact-checking websites are used to assess a claim. However, these websites require domain experts to assign a credibility value, hence, fact-checking is a time-consuming process. Moreover, several approaches have been proposed to automatically detect fake news using supervised classification models. These works extract features from news content such as the number of nouns, length of the article, fraction of positive/negative words, etc. (Hardalov et al., 2016; Rubin et al., 2016; Horne and Adali, 2017). Besides, some prior works proposed propagation-based models for evaluating news credibility (Gupta et al.; Jin et al., 2016, 2014). Nonetheless, they initialized credibility values for the entire network using a supervised classifier. In this paper, we propose a semi-supervised content-based fake news detection method which exploits tensor decomposition, k-nearest neighbor, and belief propagation algorithm. Our main contributions are:



(a) Dataset1 (Horne and Adali, 2017) (b) Dataset2 (Hardalov et al., 2016)

Figure 1: Performance of our proposed method with different percentage of labeled news on two public datasets.

- We leverage tensor-based article embeddings, which are shown to produce a concise representation of articles with respect to their spatial context, in order to derive a graph representation of news articles.
- We formulate fake news detection as a semi-supervised method that propagates known labels on a graph to determine unknown labels.
- We evaluate our method on real datasets. Experiments on two previously used datasets demonstrate that our method outperforms prior works since it requires a fewer number of known labels and achieves comparable performance.

2 Problem Formulation

Given a collection of news articles $N = \{n_1, n_2, n_3..., n_m\}$ and a vector $l \in \{-1, 0, 1\}$ with partial labels where $-1, 1$ and 0 represent fake, real and unknown news, respectively. Our goal is to classify the unlabeled news.

3 Proposed Method

Our method consists of the following steps:

Step 1: Tensor Decomposition. Given a set of news where each news article is a vector that contains its words content, we build a three-mode tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times I \times M}$ (words, words, news) where for each article, we create a co-occurrence matrix where (word, word) appear within a window of 5-10 words[1]. This tensor-based article embedding was proposed by (Hosseinimotlagh and Papalexakis, 2018) and captures spatial/contextual nuances of different types of news articles, and as (Hosseinimotlagh and Papalexakis, 2018) demonstrates, it results in homogeneous groups. Here, we use a similar tensor representation and use CP/PARAFAC tensor decomposition (Harshman, 1970) to factorize the tensors, obtaining the factor matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ whose columns correspond to different latent topics, clustering news articles and words in the latent topic space. More specifically, each row of $\mathbf{C}$ is the representation of the corresponding article in that embedding space.

Step 2: K-nearest neighbor graph of news. The tensor embedding we computed in Step 1 provides a compact and discriminative representation of articles into a concise set of latent topics. Using this embedding, we construct a k-nearest neighbor graph G using the factor matrix $\mathbf{C}$ from tensor decomposition. We compute the Euclidean distance and select the k-closest news per each news. Hence, each node in G represents a news article and each edge represents that two articles are correlated. If article n_1 is neighbor of article n_2, then we consider that n_2 is also neighbor of n_1. Thus, the resultant graph is an undirected graph which is represented in an adjacency matrix.

Step 3: Belief Propagation. Using the graphical representation of the articles above, and considering that for a small set of those nodes we have ground truth labels, our problem becomes an instant of semi-supervised learning over graphs. We use Belief Propagation assuming homophily (i.e., articles that are connected in this representation are likely to be of the same type). More specifically, we use a fast and linearized variant of Belief Propagation, the FaBP algorithm (Koutra et al., 2011).

[1]We experimented with small values of that window and the results are qualitatively similar.

4 Experimental Evaluation

We implemented our method in Matlab using Tensor Toolbox (Bader et al., 2015) and FaBP proposed in (Koutra et al., 2011).

We evaluate the performance of our method on three datasets. We collected news articles in a 3-months period from June to August, 2017. Our dataset contains 33,160 fake news and 33,047 real news. In addition, we use two datasets provided by (Horne and Adali, 2017) and (Hardalov et al., 2016). One dataset consists of 150 political news, 75 news for each class (fake/real), and the another dataset contains 68 real news an 69 fake news.

We performed an iterative process using cross-validation where we tried different settings respect to decomposition rank and number of neighbors. We found that the best accuracy is obtained when those parameters are set to 10. In all experiments, we tested accuracy on the news whose labels were unknown. We evaluated our method using different percentages of known labels to show that we can perform better than existing models only using few labeled news (Figure 1). We achieve 72% accuracy using Dataset1 with only 20% of news labels while Horne and Adali (2017) achieved 71% accuracy. For Dataset2, we achieve 67.38% accuracy only using 10% of labels while Hardalov et al. (2016) achieved 61.31% accuracy using linguistic features and 70% data for training their model. Additionally, we evaluate our method using our dataset, we achieve 70.01% accuracy with 10% of news labels. It should be noted our method achieves good performance only using few labeled news because our k-nearest neighbor graph captures correlation among news based on their content which is exploited during belief propagation to distinguish fake from real news.

5 Conclusions

In this paper, we propose a semi-supervised content-based method for fake news detection which leverages tensor-based article embeddings. Experiments on three real-world datasets show our model is able to distinguish fake from real news only using a small number of labeled news, compared to state-of-the-art content-based approaches who achieve similar quality while assuming a fully supervised model.

References

Brett W. Bader, Tamara G. Kolda, Jimeng Sun, Evrim Acar, and et al. 2015. Matlab tensor toolbox version 2.6. Available online.

Manish Gupta, Peixiang Zhao, and Jiawei Han. *Evaluating Event Credibility on Twitter*.

Momchil Hardalov, Ivan Koychev, and Preslav Nakov. 2016. In search of credible news. *Artificial Intelligence: Methodology, Systems, and Applications. AIMSA 2016 Lecture Notes in Computer Science*, page 172180.

R. A. Harshman. 1970. Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 16(1):84.

Benjamin D. Horne and Sibel Adali. 2017. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. *CoRR*, abs/1703.09398.

Seyedmehdi Hosseinimotlagh and Evangelos E Papalexakis. 2018. Unsupervised content-based identification of fake news articles with tensor decomposition ensembles. In *WSDM 2018 Workshop on Misinformation and Misbehavior Mining on the Web (MIS2)*. ACM.

Z. Jin, J. Cao, Y. G. Jiang, and Y. Zhang. 2014. News credibility evaluation on microblog with a hierarchical propagation model. In *2014 IEEE International Conference on Data Mining*, pages 230–239.

Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. 2016. News verification by exploiting conflicting social viewpoints in microblogs.

Danai Koutra, Tai-You Ke, U. Kang, Duen Chau, Hsing-Kuo Pao, and Christos Faloutsos. 2011. Unifying Guilt-by-Association Approaches: Theorems and Fast Algorithms. In *Machine Learning and Knowledge Discovery in Databases (ECML/PKDD)*, volume 6912 of *Lecture Notes in Computer Science*, pages 245–260.

Victoria L. Rubin, Niall J. Conroy, Yimin Chen, and Sarah Cornwell. 2016. Fake news or truth? using satirical cues to detect potentially misleading news.