Aiding Sentiment Evaluation with Social Network

Pengfei Gao, Fan Yang, Hao Yin *
Institute for Computational and Mathematical Engineering
Stanford University
Stanford, CA 94305
{pfgao, fanfyang, yinh}@stanford.edu

Abstract

Some abstract

1 Introduction

Sentiment analysis is one of the important task in natural language processing community[8] which help people navigate the huge amount of user-generated content available online. Machine learning systems that make decision on the attitude of viewpoints to be positive, neutral or negative that enable people to understand the enormous body of opinions on the Internet, ranging from product reviews to political positions.

Interestingly, most of the viewpoints nowadays can be obtained from online website that actually has a social network behind it, since user-generated content often appears in the context of social media. Therefore nowadays user-relationship information is now more easily obtainable. For example, huge amount of tweets from Twitter express people's opinions on different subjects. Each tweet is associated with a user and users formed social network structure through the mechanisms of "follower". When a user forms a link in the network such as Twitter, they tend to have a personal relationship then the principle in language called "homophily" suggests that users who are connected via some social relationship may also share similar opinions or linguistic variation. Figure 1 from [1] gives an example of how users from different communities may understand the word "sick" differently.

In this paper, we are going to explore different methods that utilize social network information in sentiment analysis with deep learning. Network structure is useful and informative in NLP-related task as people within each community may have their own "jargon" in expressing ideas and sentiments.

2 Related Work

Sentiment Analysis

The current state-of-the-art method is convolutional neural networks(CNN)[9] which takes word embeddings from sentences as inputs and output a softmax classification to identify sentence sentiment. The typical structure of such CNN is some convolutional layer on top of original sentence word embeddings, then a max pooling layer on top of the convolutional layer to extract some extreme information. Finally, a dense layer with fully connected network is added to transform features from CNN to a softmax classifier. A simple CNN model with one convolutional layer of two-width window plus one max pooling layer with single channel can achieve amazing result[1]. Different

*Each member contributes equally, and names are put in alphabetic order.

initializations methods[11] could also be adopted to improve prediction accuracy, but they all share the similar structure as described above.

Network Node Embeddings

Aiding Classification with Social Network

The intuition behind combing social network with classification is that users connected are more likely to hold similar opinions and use language similarly.

Tan et al.(2011)[5] is the first paper to show social relationship information can be exploited to improve sentiment analysis. They have shown numerically that incorporating social-network information can indeed lead to statistically significant sentiment-classification improvements over the performance of a SVM baseline model that only has access to textual features. Yang and Eisenstein(2016)[1] is a more recent version for combining social network information and sentiment analysis. They study task for classifying sentiment to be positive, neutral and negative for each tweets given text and user ID information. Their model consider the author information and sentence information separately: each node (author) in the network is assigned an embedding vector using the LINE algorithm[2], and then is (softly) assigned each cluster on the network based on the embedding. Each cluster has its own model, which is a CNN model combined with max pool layer.

On the other hand, Yang and Chang(2016)[6] study another problems called entity linking, which is the task of identifying mentions of entities in text, and linking them to entries in a knowledge base. They achieve the-state-of-art result with a tree-based model in Twitter data. To further improve the performance, Yang et al.(2016)[7] propose to incorporate social network information in the same problem. Intuitively, socially linked individuals share interests, and are therefore likely to mention the same sorts of entities. They build a bilinear model based on the previous the-state-of-art tree model[6] that consider interactions of users and entities. This new model incorporating social network information has a F1 improvements of 1%-5% on benchmark datasets.

3 Problem Definition and Data Description

We obtained the data used by [?]. The data consists of a collection of tweets as well as some network information on Twitter. For each tweet sample, we have one tweet ID, one user ID, a sentiment label (positive, neutral, negative), and the tweet content itself. For example, the following are two examples of our data samples.

```
261140278944088066 17572408 negative @USER may i have an industrial revolution ...
237571817550786563 727519172 neutral @USER i told you shane would get his 5th-star .
```

We have three networks, i.e., FOLLOWER, MENTION and RETWEET network which have been explained in details by [?].

4 Methodology

Our project will start from the method proposed by [?], which is summarized in Figure 1 (black lines). The model consider the author information and sentence information separately: each node (author) in the network is assigned an embedding vector using the LINE algorithm ([?]), and then is (softly) assigned each cluster on the network based on the embedding. Each cluster has its own model, which is a CNN model combined with max pool layer. This model provides a baseline in our project.

We are going to extend this model in three aspects, as is illustrated in the red lines in Figure 1. To be specific,

(1) Using other node embedding methods in network analysis, such as *DeepWalk* ([?]) and *node2vec* ([?]);

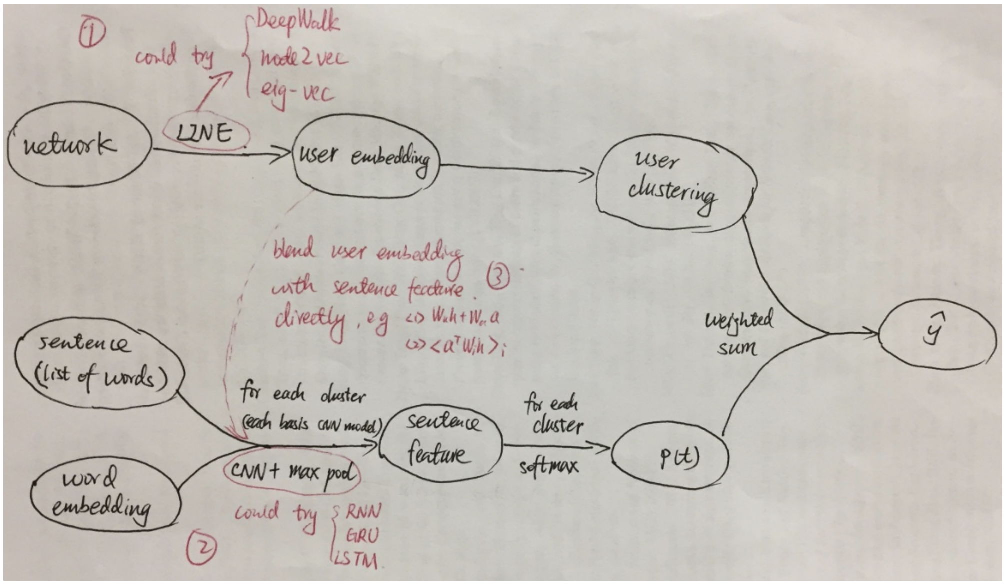

Figure 1: General Methodology

(2) Explore other methods to combine author information and sentence information, especially the bilinear form $a^T W h$ which measures the interaction between author and sentence. Here a is the author embedding and h is the sentence embedding.

(3) Explore other models in sentiment analysis, such as RNN, GRU, and LSTM.

5 Experiments

Embedding method	CNN	DeepWalk	LINE	node2vec	random
Dev2013	68.85	67.71	**69.51**	68.58	68.50
Test2013	69.53	67.58	**69.67**	68.58	68.49
Test2014	**72.41**	71.46	71.44	71.46	71.69
Test2015	64.40	**64.71**	64.57	64.25	63.50
Avg test sets	**68.78**	67.92	68.56	68.10	67.89

Table 1: Prediction performance on each Dev and Test Sets.

6 Discussion

Reference

[1] Yang, Yi, and Jacob Eisenstein. ”Overcoming Language Variation in Sentiment Analysis with Social Attention.” arXiv preprint arXiv:1511.06052 (2016).

[2] Perozzi, Bryan, Rami Al-Rfou, and Steven Skiena. ”Deepwalk: Online learning of social representations.” Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2014.

[3] Tang, Jian, et al. ”Line: Large-scale information network embedding.” Proceedings of the 24th International Conference on World Wide Web. ACM, 2015.

[4] Grover, Aditya, and Jure Leskovec. ”node2vec: Scalable feature learning for networks.” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016.

[5] Tan, Chenhao, et al. ”User-level sentiment analysis incorporating social networks.” Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2011.

[6] Yang, Yi, and Ming-Wei Chang. ”S-mart: Novel tree-based structured learning algorithms applied to tweet entity linking.” arXiv preprint arXiv:1609.08075 (2016).

[7] Yang, Yi, Ming-Wei Chang, and Jacob Eisenstein. ”Toward socially-infused information extraction: Embedding authors, mentions, and entities.” arXiv preprint arXiv:1609.08084 (2016).

[8] Pang, Bo, and Lillian Lee. ”Opinion mining and sentiment analysis.” Foundations and Trends in Information Retrieval 2.12 (2008): 1-135.

[9] Kim, Yoon. ”Convolutional neural networks for sentence classification.” arXiv preprint arXiv:1408.5882 (2014).

[10] Dos Santos, Ccero Nogueira, and Maira Gatti. ”Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts.” COLING. 2014.

[11] Severyn, Aliaksei, and Alessandro Moschitti. ”Twitter sentiment analysis with deep convolutional neural networks.” Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2015.