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

In this project, 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

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,

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

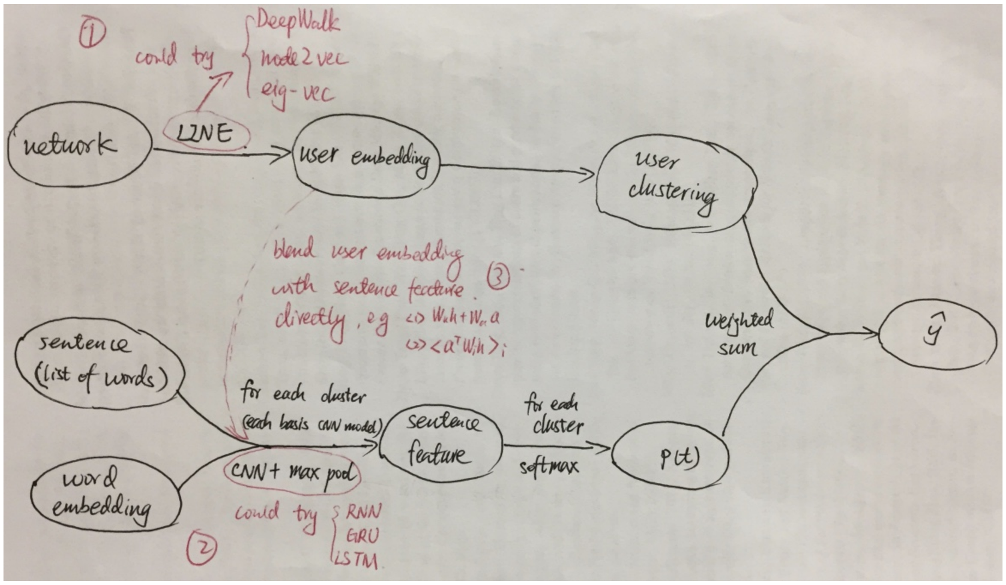

Figure 1: General Methodology

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

(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).