
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

Milestone report is seen in Section 2. For reader’s reference, we give an introduction of our project in this section.

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 [Yang and Eisenstein(2017)]. 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 [Yang and Eisenstein(2017)].

4 Methodology

Our project will start from the method proposed by [Yang and Eisenstein(2017)], 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 ([Tang et al.(2015)Tang, Qu, Wang, Zhang, Yan, and Mei]), 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.

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

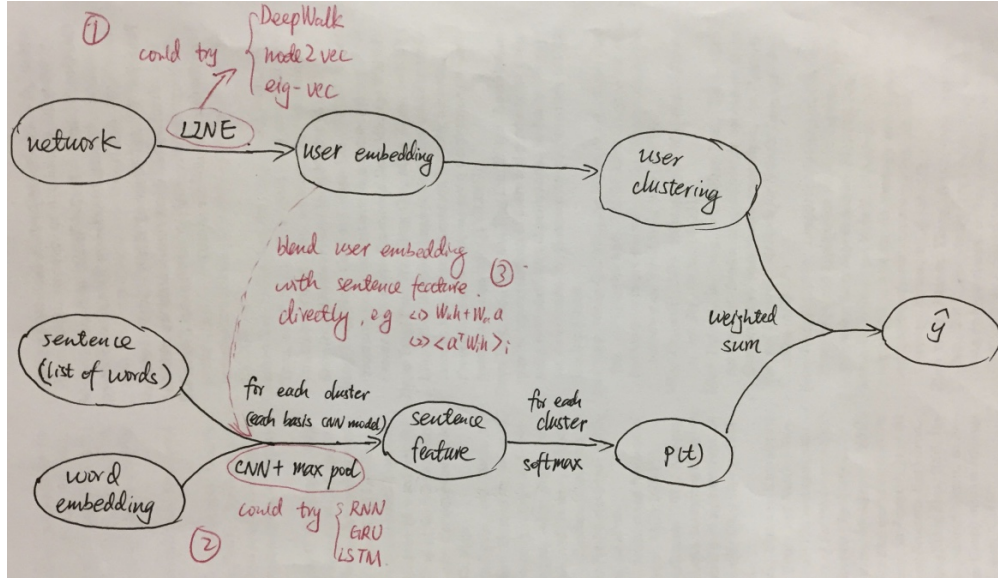


Figure 1: General Methodology

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* ([Perozzi et al.(2014)Perozzi, Al-Rfou, and Skiena]) and *node2vec* ([Grover and Leskovec(2016)]);
- (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