Aiding Sentiment Analysis with Social Network

Pengfei Gao, Fan Yang, Hao Yin

Background

Intuition: the tendency of socially linked individuals to use language in similar ways.

Our Goal: exploit social networks information to make sentiment analysis adapt to social language variation.

Previous Approaches: CNN is state-of-the-art approach, but it solely exploits texts information.

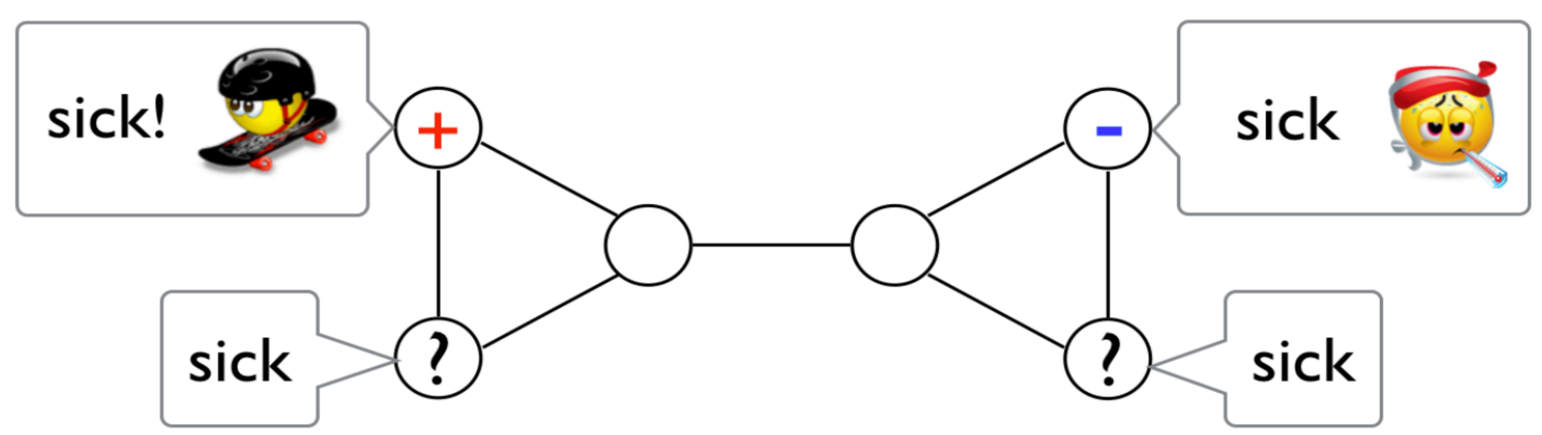


Figure 1: Words such as 'sick' can express opposite sentiment polarities depending on the author. We account for this variation by generalizing across the social network.

Problems & Dataset

SemEval Twitter sentiment analysis tasks

Goal

classify each message as positive, negative, or neutral.

Train Set(min Cross Entropy)

SemEval Twitter 2013 includes tweets and user ID

Evaluation Set (max F1 score)

SemEval Twitter 2013–2015 includes test tweets and user ID

Network	# Author	# Relation
FOLLOWER+	18,281	1,287,260
MENTION+	25,007	1,403,369
RETWEET+	35,376	2,194,319

Social Networks

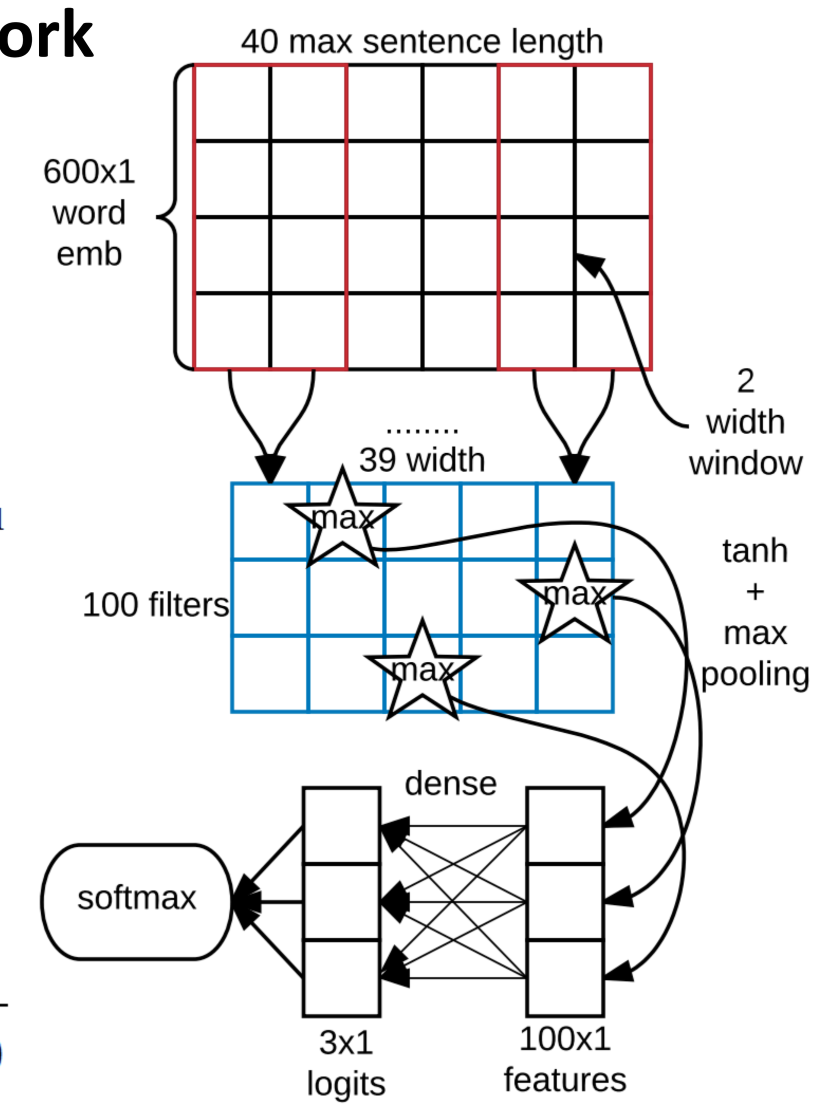
- follow
- mention
- retweet

Dataset	# Positive	# Negative	# Neutral	# Tweet
Train 2013	3,230	1,265	4,109	8,604
Dev 2013	477	273	614	1,364
Test 2013	1,572	601	1,640	3,813
Test 2014	982	202	669	1,853
Test 2015	1,038	365	987	2,390

Methods

Convolutional Neural Network

CNN is the state-of-the-art methods for sentiment analysis NLP tasks.



Our network structure

(0) Input $h_i \in \mathbb{R}^{600\times1}, i = 1:40$

(1) one convolutional layer

$$c_i = \tanh(W_L h_i + W_R h_{i+1} + b) \in \mathbb{R}^{100\times1}$$

(2) one max pooling layer

$$s = \max_{i=1:39} c_i \in \mathbb{R}^{100\times1}$$

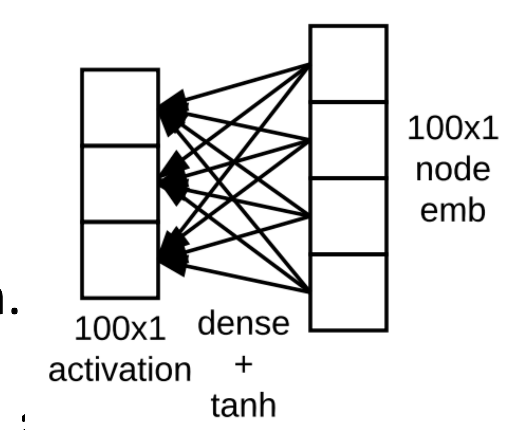
(3) finally a logistic regression

$$P(\mathrm{Y}=\mathrm{c} \mid \mathrm{s}) = \frac{\exp(\beta_c^T s + b_c)}{\sum_{c'} \exp(\beta_{c'}^T s + b_{c'})}$$

Network Node Embedding

- DeepWalk, LINE, Node2vec

These algorithms learn network node embeddings from network edges information.



Merge two models via element-wise

(1) Apply a dense network to convert author embeddings to the same dimension of convolution filters.

$$a \in \mathbb{R}^{100\times1}, z = \tanh(W_a a + b_a) \in \mathbb{R}^{100\times1}$$

(2) Use this new author-activated convolution layer for the rest procedures

$$\tilde{c}_i = z \circ c_i, \tilde{s} = \max_{i=1:39} \tilde{c}_i \in \mathbb{R}^{100\times1}, P(\mathrm{Y}=\mathrm{c} \mid \mathrm{s}) = \frac{\exp(\beta_c^T \tilde{s} + b_c)}{\sum_{c'} \exp(\beta_{c'}^T \tilde{s} + b_{c'})}$$

Note: this activation is equivalent to bilinear form that models *interaction between author embeddings and word embeddings*.

Hyper parameters

Dropout rate 0.4 in CNN; L2 Penalty 0.01 in all kernel weights

Results

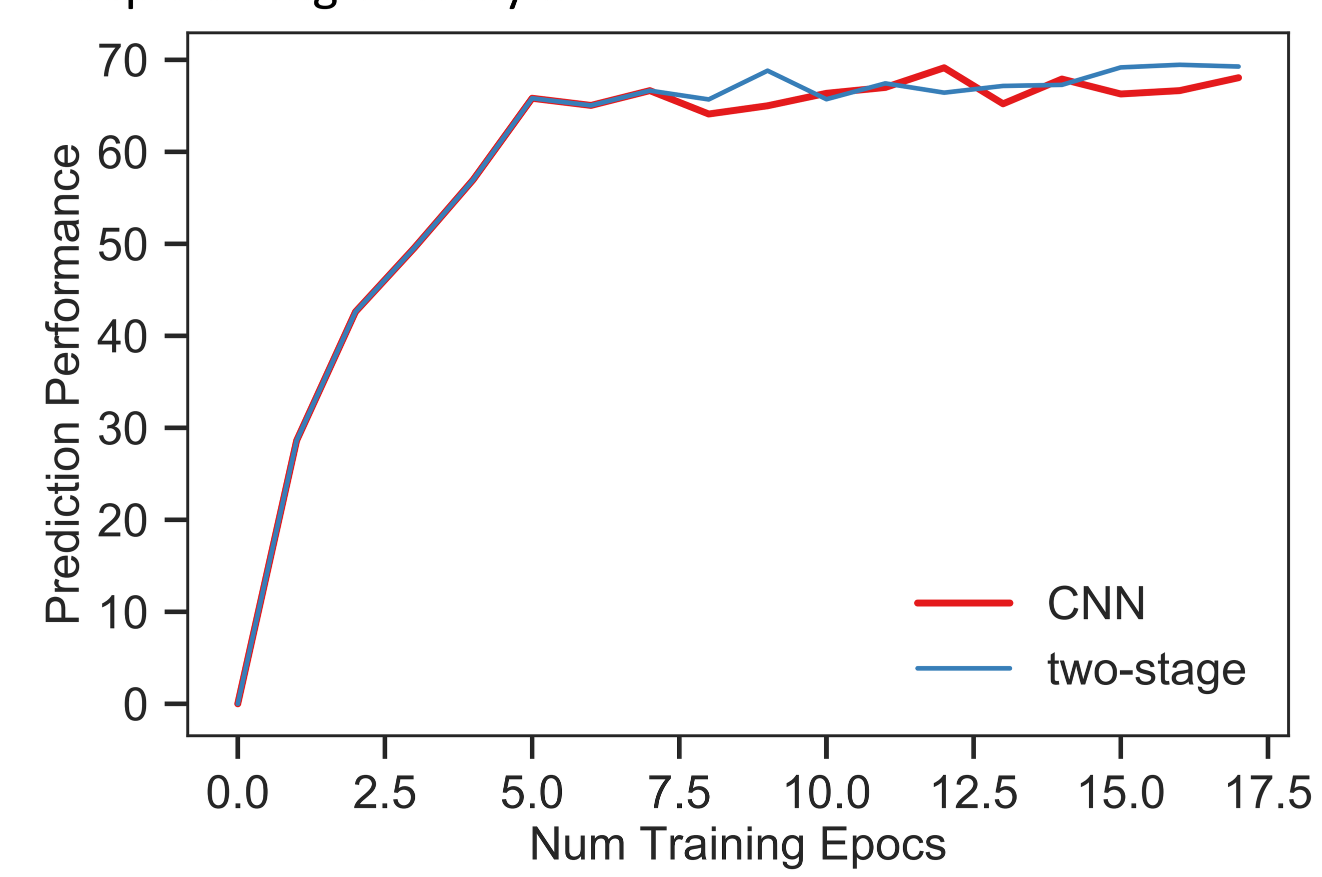
Sentence CNN vs Author Interaction

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

Two-stage Training

- Adding author embedding only after several epochs of CNN training

1) fix CNN layer
2) keep training CNN layer



Discussions

- Our experiments suffer serious overfitting, which maybe solved with bigger datasets.
- Current author embedding seems non-informative, which might due to incompleteness of networks (need a better network than twitter).
- More models may be explored, such as RNN, GRU, LSTM.
- Despite the frustrating experiment results, the idea of tracking langrage variation with network is still promising.

Reference

- Y. Yang and J. Eisenstein. Overcoming Language Variation in Sentiment Analysis with Social Attention. TACL 2017.
- B. Perozzi, R. Al-Rfou, and S. Skiena. DeepWalk: Online learning of social representations. KDD 2014.
- J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei. LINE: Large-scale Information Network Embedding. WWW 2015.
- A. Grover and J. Leskovec. node2vec: Scalable Feature Learning for Networks. KDD 2016.