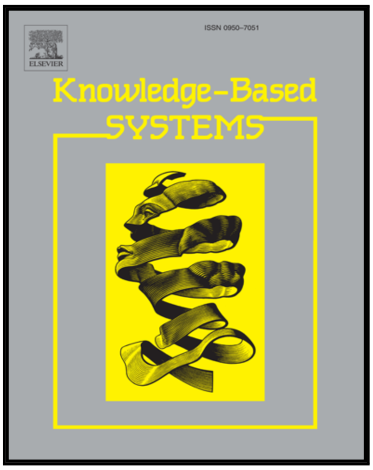
Accepted Manuscript



Microblog Sentiment Analysis with Weak Dependency Connections

Zou Xiaomei, Yang Jing, Zhang Jianpei, Han Hongyu

PII: S0950-7051(17)30566-X
DOI: 10.1016/j.knosys.2017.11.035
Reference: KNOSYS 4130

To appear in: *Knowledge-Based Systems*

Received date: 16 May 2017
Revised date: 26 November 2017
Accepted date: 27 November 2017

Please cite this article as: Zou Xiaomei, Yang Jing, Zhang Jianpei, Han Hongyu, Microblog Sentiment Analysis with Weak Dependency Connections, *Knowledge-Based Systems* (2017), doi: 10.1016/j.knosys.2017.11.035

Microblog Sentiment Analysis with Weak Dependency Connections

Zou Xiaomei[a], Yang Jing[a,*], Zhang Jianpei[a], Han Hongyu[a]

[a] *College of Computer Science and Technology, Harbin Engineering University, Heilongjiang, 150001, China*

Abstract

With the rise of microblogging services like Twitter and Sina Weibo, users are able to post their real-time mood and opinions conveniently and swiftly. At the same time, the ubiquitous social media results in abundant social relations such as following and follower relations. Social relations create a new source for microblog sentiment analysis, which attracts a great amount of attention in recent years. There are two theories that support the use of social relations for sentiment analysis - sentiment consistency and emotional contagion. However, most existing microblog sentiment analysis methods only employ direct connections which cannot fully use the heterogeneous connections in social media. As online social networks consist of communities and nodes in the same community which form weak dependency connections usually share similarities, we investigate how to exploit weak dependency connections as an aspect of social contexts for microblog sentiment analysis in this paper. In particular, we employ community detection methods to capture

*Corresponding author.

Email addresses: zouxiaomeihy@163.com (Zou Xiaomei), yangjing@hrbeu.edu.cn (Yang Jing), zhangjianpei@hrbeu.edu.cn (Zhang Jianpei), hanhongyu@hrbeu.edu.cn (Han Hongyu)

weak dependency connections and propose a new model for microblog sentiment analysis which incorporates weak dependency connections, sentiment consistency, and emotional contagion together with text information. Experimental results on two real Twitter datasets demonstrate that our proposed model can outperform baseline methods consistently and significantly.

Keywords: Sentiment Analysis, Microblog, Social Relation, Community Detection, Twitter

1. Introduction

With the development of Web 2.0, a series of new sites such as social networking sites, microblogging and forums have emerged, users become the main contributors to Internet content. Users use these new sites as a platform to exchange their views and feelings, so this kind of platform contains a volume of user-generated content with various types such as text and images. It is of great value and has a wide range of application prospects such as customer relationship management, recommendation systems, and business intelligence [3, 41, 42] to study how to obtain the true sentiment of users from user-generated information with different lengths and various expressions. However, user-generated information is unstructured, hence the machine cannot process it directly. In particular, for text information, its automatic analysis requires a deep understanding of natural language by machines, which is a great challenge we are facing [43].

Because of its great value, sentiment analysis has attracted a large number of researchers to study it. They put forward a variety of methods, the two main mostly used methods are the lexicon-based method [47, 48] and the method based on machine learning [44–46, 51]. There are also some scholars that try to combine the two methods together. These traditional methods usually only use text content as features to analyze microblog sentiments. They assume that texts are independent of each other. However, this assumption is clearly not in line with new sites such as microblogging platform for microblogs are networked data. There exists a variety of connections between them which we call contexts in this paper, such as semantic connections and so on. In addition, in view of the different backgrounds and standpoints of users, their views vary a lot on the same problem. For example, location is one reason for people holding different views on the issue of Terminal High Altitude Area Defense (THAAD) deployment in South Korea. People in Seongju County, South Korea, fear that the system has a negative impact on their health, so they are opposed to the THAAD deployment. While most of the rest people in South Korea are in favor of the THAAD deployment.

To improve microblog sentiment analysis and based on the assumption that microblogs are not independent of each other, a natural idea is to use the links between microblogs to assist in sentiment analysis. Many scholars have studied this problem, such as [34] and [16]. These methods usually use the following and follower relations between users as social contexts to aid sentiment analysis. The theories behind using social relations in sentiment analysis are sentiment consistency [1] and emotional contagion [15]. Sentiment consistency, which is also called user contexts and is regarded as an aspect of social contexts, indicates microblogs posted by the same person tend to have the same sentiment label. Emotional contagion implies similar users tend to have the same opinion, so microblogs posted by them may also share the same sentiment label. Emotional contagion is usually called user relation contexts, which is also an aspect of social contexts. The phenomenon of emotional contagion is called homophily [39], it is also known as "birds of a feather flock together". However, traditional methods only use direct social relations between microblogs or users to get homogeneous connections, ignoring the influence of indirect connections on sentiment analysis.

Community structure is a common characteristic of networks. Connections between communities are relatively sparse, while connections within communities are relatively dense. Nodes in the same community usually share some characteristics, even if they are not directly connected. That is, there is weak dependency connections between these nodes [37] [1]. For example, the THAAD deployment issue, we can divide people into two communities: people in Seongju County and people not in Seongju County. Weak dependency connections can provide crucial context information about users' interests [35], which have proven to be useful in job hunting [13], the diffusion of ideas [14], knowledge transfer [21] and relational learning [37], while they are rarely explored in sentiment analysis. Therefore, in this paper, we treat weak dependency connections also as an aspect of social contexts which we call "Weak Dependency Connection Contexts". In online social networks, microblogs also form a network for microblogs are posted by users and users are connected with each other. So we build the microblog network by using user-microblog relations and user relations. Then this network is used to

[1] In this paper, week dependency connections represent edges within communities, which are different from weak ties (connections that bridge communities).

extract weak dependency connections between microblogs. In this paper, we are going to study whether weak dependency connections between microblogs can affect sentiment analysis or not and try to combine sentiment consistency (user contexts), emotional contagion (user relation contexts), and weak dependency connections (weak dependency connection contexts) together with text information into microblog sentiment analysis.

The main contributions of this paper include:

1. Using user contexts and user relation contexts to acquire weak dependency connections between microblogs;
2. Providing a principled way to model weak dependency connections for microblog sentiment analysis;
3. Proposing a novel microblog sentiment analysis model which incorporates weak dependency connections, sentiment consistency and emotional contagion; and
4. Evaluating the proposed model extensively using real-world datasets to understand the working of the proposed model.

The remainder of this paper is organized as follows. In Section 2, some related works about microblog sentiment analysis and community detection are introduced. In Section 3, we define the problem we study and propose our model. In Section 4, the experimental results are presented. In Section 5, we conclude the whole paper.

2. Related Work

2.1. Microblog Sentiment Analysis

Microblog sentiment analysis has become a hot research topic in these years [22, 27, 34]. Because microblogs are short and noisy, sentiment analysis of microblogs is more challenging. Many methods are proposed to solve this problem. Go et al. [12] used emoticons as distant supervision features to analyze the sentiment of tweets. They compared different machine learning methods on microblog sentiment analysis. In [6], generalized emoticons, repeated punctuations, and repeated words were used to build a co-occurrence graph by label propagation algorithm and the co-occurrence graph was used to identify the sentiment polarities of tweets. Kiritchenko et al. [18] built a sentiment lexicon using the relations between words and emoticons, then they used the lexicon to extract sentiment features and analyze microblogs. All these methods mentioned above utilize text information only and ignore the extra information provided by the microblog media.

In recent years, there are more and more research works about how to utilize user information to analyze sentiment. Tan et al. [34] proposed a method using user follower/followee relations and "@" information to identify the sentiment of users on Twitter. Wu et al. [40] took sentiment analysis of users to a specific topic as a problem of collaborative filtering, relations between users were applied to predict sentiment of users. Similarly, Speriosu et al. [33] also exploited user relations graph. The classification results of the maximum entropy model were used as labels and then the authors implemented label propagation algorithm to identify sentiment. West et al. [38] built a model using the signed social network to predict individual A's opinion of individual B. Fu et al. [10] created a following graph and a follower graph respectively, they added the two graphs into the Bayesian and SVM classifier so as to improve the Bayesian and SVM classifier accuracy rate on the task of microblog sentiment analysis. Fersini et al. [7] proposed Approval Network as a novel graph representation to jointly model sentiment consistency [1] and emotional contagion [15]. They applied the new network into user-level sentiment analysis and aspect-level sentiment analysis. Cheng et al. [4] refined the user relations. In their paper, the user relations were divided into positive and negative relations, an unsupervised method was utilized to user-level sentiment analysis.

These works are user-level or user-topic level sentiment classification methods, while our method is microblog-level. Hu et al. [16] proposed a framework named SANT (a **S**ociological **A**pproach to handling **N**oisy and short **T**exts) combining social contexts to classify sentiment of microblogs. On the basis of [16], Lu [23] added content similarity to the framework of SANT and proposed a semi-supervised method to identify sentiment of tweets. Wu et al. [40] argued the framework proposed by Hu et al. [16] was a purely content-based approach, so they proposed a **S**tructured **M**icroblog **S**entiment **C**lassification (SMSC) framework which used social contexts at the prediction stage.

The disadvantage of these methods is that these methods only consider the direct social relations between users or microblogs. According to the observation of the previous section, even if there are no direct relationships between users, they may still have the same view. So we propose to use weak dependency connections to aid microblog sentiment analysis.

2.2. Community Detection

Over the past few decades, the rapid development of new technologies and the recent globalization of the commercial environment have brought human beings into the Information Age. A new area of research named "network science" has attracted lots scholars. The theoretical basis of the new field arises from graph theory, statistical and probability theory, social structure as well as data mining. With an in-depth study of networks, it has been discovered that community structure is a common characteristic of networks, also known as clustering, i.e. the organization of vertices in clusters [52]. There are many edges joining vertices within the same clusters, while few edges joining vertices between different clusters. Detecting communities is of great importance and has a wide range of application prospects in sociology, biology and computer science, disciplines where systems are often represented as graphs. For example, it can be used to improve the performance of services provided on the World Wide Web [53], to build efficient recommendation systems [36, 54], to search paths [55] and so on.

Community detection algorithms are divided into two types, non-overlapping community detection methods, and overlapping community detection methods. In this paper, non-overlapping community detection algorithms are adopted because the microblog polarity is either positive or negative, i.e., the microblog polarity cannot be both positive and negative at the same time. Non-overlapping community detection methods commonly used can be divided into: community detection algorithms based on modularity optimization [5, 11, 24], community detection algorithms based on information theory [30, 31], community detection algorithms based on label propagation [20, 29], community detection algorithms based on graph theory [8, 25, 26, 28, 59].

The idea of community detection algorithms based on modularity optimization is to define the problem of community detection as an optimization problem, and then search the community structure with the optimal target value. The modularity Q value first proposed by Newman [24] is currently the most widely used optimization target. The index measures the significance of the community structure by comparing the difference between the edge density of each community in the real network and the edge density of the corresponding subgraph in the random network.

Community detection algorithms based on information theory are drawn from the perspective of information theory. These methods take the modular description of the network as a lossy compression of the network topology [30, 31]. So they transform the problem of community detection into a fundamental problem in information theory: finding an effective way to compress the topology.

The idea of the community detection algorithms based on label propagation is that the edges of a complex network represent the propagation of information between individuals, and the result of propagation is that nodes within the community share the same information. Raghavan et al. [29] proposed a fast label propagation algorithm based on this idea (referred to as LPA algorithm). The LPA algorithm first assigns a unique label to each node. In each iteration, each node updates its own label to the one with the largest number of occurrences of its neighbors, and if there are multiple labels, a random selection is made as an update value. After a number of iterations, the densely connected nodes will converge on the same label, and ultimately, nodes with the same label are clustered into a community.

In this paper, to fully investigate whether weak dependencies can influence microblog sentiment analysis or not, four classic community detection algorithms (Louvain [2], Infomap [31], LPA [29], walktrap [28]) are used to detect weak dependencies between microblogs. They are from the above four types of methods respectively and are always used as baselines in community detection.

3. Model

The proposed model in this paper is shown in Figure 1. In this model, first, we use user contexts and user relation contexts to build a direct relation graph of microblogs. Second, we use community detection algorithms to get weak dependency connections between microblogs according to the direct relation graph. User contexts, user relation contexts and weak dependency connection contexts are called social contexts. At last, we combine user contexts, user relation contexts, and weak dependency connection contexts together. We join the three social contexts (social contexts matrix A in Figure 1) and text information (feature matrix X in Figure 1) into the final sentiment classification model. We will introduce every step in detail in this section.

3.1. Notation

In this paper, uppercase letters like B are used to denote matrixes, lowercase bold letters like $\mathbf{x}$ denote vectors. and lowercase letters like a denote numbers. We use B_{i*} to denote the i-th row of matrix B and B_{*j} to denote the j-th column of matrix B. The entry at the i-th row and j-th column is

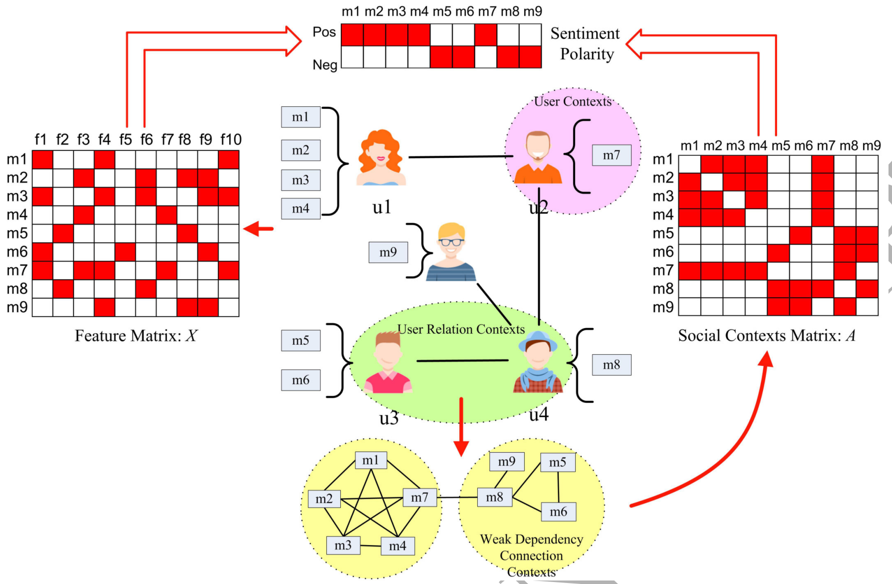


Figure 1: The Microblog Sentiment Analysis Model using Weak Dependency Connections

denoted as B_{ij}. B^T is used as the transposition of matrix B. $\|B\|_F$ denotes the Frobenius norm of B. $tr(.)$ is the trace of a matrix.

The goal of this paper is by using the training set feature matrix $X \in R^{n\times m}$ (where n represents the number of microblogs in training set, m represents the number of features.) and label matrix $Y \in R^{n\times c}$ (where c is the number of sentiment polarities) to construct a classifier $W \in R^{m\times c}$, and then classifier $W \in R^{m\times c}$ is used to predict the sentiment of an unseen microblog $\mathbf{x}$. Y represents ground truth labels of microblogs. We use $\hat{Y} = XW \in R^{n\times c}$ to represent the fitted values of the ground truth label matrix Y. In particular, elements in feature matrix X are unigrams, $X_{ij} = 1$ if and only if the i-th microblog contain the j-th unigram. In this paper, we only consider the binary classification of sentiments, that is, $c = 2$. Therefore, if a microblog is positive, then its ground truth label is $Y_{i*} = [+1 \quad -1]$. And if the sentiment of a microblog is negative, then its label is $Y_{i*} = [-1 \quad +1]$. $U \in R^{d \times n}$ is a user-microblog matrix where $U_{ij} = 1$ if the i-th user posts the j-th microblog. d is the number of users. We use $F \in R^{d \times d}$ to represent a user-user matrix where $F_{ij} = 1$ if there exists a following/followee relation between the i-th user and the j-th user.

Given an undirected graph $G = (V, E)$, A represents its adjacency matrix, $L = D - A$ represents the Laplacian matrix of G [57], where D is diagonal matrix and D_{ii} indicates the degree of the i-th vertex.

To classify an unseen microblog, we use the prediction function in Equation 1. All the variables and their meanings are shown in Table 1.

$$g(\mathbf{x}) = \begin{cases} +1 \quad if \quad \mathbf{x}W_{*1} > \mathbf{x}W_{*2} \\ -1 \quad if \quad \mathbf{x}W_{*1} < \mathbf{x}W_{*2} \\ +1 \quad or \quad -1 \quad randomly \quad if \quad \mathbf{x}W_{*1} = \mathbf{x}W_{*2} \end{cases} \tag{1}$$

3.2. Datasets

In this paper, our experiments are conducted on two Twitter sentiment analysis benchmark datasets: HCR and OMD. Many proposed works use these two datasets to evaluate the performance of using social relations for sentiment analysis. These two datasets include raw texts and sentiment labels labeled by manual.

HCR: This dataset is collected by Speriosu et al. [33]. It includes tweets about health care reform of America in March 2010. It has three parts: training set, development set, and test set. There are five kinds of labels in

Table 1: Meaning of Variables

Variables	Meaning	Type
X	feature matrix	$R^{n\times m}$
Y	ground truth label matrix	$R^{n\times c}$
$\hat{Y}$	fitted sentiment label matrix	$R^{n\times c}$
n	number of features	integer
m	number of training set	integer
c	number of sentiment classification	integer
W	classifier	$R^{m\times c}$
$\mathbf{x}$	feature vector of a microblog	R^{m}
U	user-microblog matrix	$R^{d\times n}$
d	number of users	integer
F	user-user matrix	$R^{d\times d}$
A	microblog-microblog relation matrix	$R^{n\times n}$
D	diagonal matrix	$R^{n\times n}$
L	Laplacian matrix	$R^{n\times n}$

the dataset: positive, negative, neutral, irrelevant and unsure. This corpus is manually annotated by the authors. In this paper, we only use tweets with positive and negative labels. We use the complete follower graph built by Kwak et al. [19] in 2009 to construct the user relations of HCR and take the graph as undirected. The dataset has 9 different topics, i.e. health care reform, Obama, Republicans, Democrats, conservatives, liberals, Tea Party, Stupak and other.

OMD: This dataset is built by Shamma et al. [32]. It consists of tweets discussing the US Presidential Debates between Barack Obama and John McCain. This dataset is manually labeled by Amazon Mechanical Turk. Every tweet is tagged by at least three Turkers and its inter-annotator agreement is 0.655 reported by Shamma et al. [32], which shows a relatively good agreement between annotators. Four kinds of labels appear in the dataset,

Table 2: Statistics of Datasets

Emoticon	HCR	OMD
# of Tweets	1434	1184
# of users	806	636
# positive Tweets	387	475
Average Tweets per User	1.78	1.86
Average Friends per User	14.95	5.54

they are positive, negative, mixed and irrelevant. We use majority voting to determine the final label of each tweet. The same as HCR, we only use tweets with the positive or negative label. The relation graph is also built by using the follower graph constructed by Kwak et al. [19] in 2009.

In this paper, we reserve microblogs posted by users that have friends and delete those microblogs whose author has no friends. The information about the two datasets is shown in Table 2.

3.3. Social Contexts

In this section, we will introduce three different social contexts (user contexts, user relation contexts and weak dependency connection contexts) and formalize them into an integrated model.

User Contexts

User contexts are based on a sociological theory called sentiment consistency. Reasonably, if two microblogs are posted by the same person, their sentiment polarities have a higher probability to be same. $A_{sc} \in R^{n \times n}$ represents the microblog-microblog matrix for sentiment consistency. We can use Equation 2 to calculate A_{sc}. $A_{scij} = 1$ if and only if the i-th microblog and the j-th microblog are posted by the same user.

$$A_{sc} = U^T \times U \tag{2}$$

Based on the sentiment consistency theory, the basic idea to integrate user contexts in sentiment classification is to make two microblogs as close as possible if they are posted by the same user. Under this situation, it can be mathematically formulated as solving the following objective function (Equation 3). In this paper, we use the Laplacian regularization method [56].

$$\begin{aligned}
&\min_W \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{sc_{ij}} \left\| \hat{Y}_{i*} - \hat{Y}_{j*} \right\|^2 \\
&= \min_W \sum_{k=1}^{c} \hat{Y}_{*k}^T (D_{sc} - A_{sc}) \hat{Y}_{*k} \\
&= \min_W tr(W^T X^T L_{sc} X W)
\end{aligned} \tag{3}$$

where D_{sc} is a diagonal matrix, $D_{sc_{ii}} = \sum_{j=1}^{n} A_{sc_{ij}}$, $L_{sc} = D_{sc} - A_{sc}$ is called the Laplacian matrix [57].

User Relation Contexts

This part is also based on a basic theory of sociology: emotional contagion. If the two users have a direct link between each other, then their views are more likely to be consistent. So we build a latent connection to make two microblogs as close as possible if they are posted by two users who are connected by a follower/friend relation. $A_{ec} \in R^{n \times n}$ is used to represent the emotional contagion matrix of microblogs. We can use Equation 4 to calculate A_{ec}.

$$A_{ec} = U^T \times F \times U \tag{4}$$

Based on the discussion above, we use Equation 5 which mathematically formulates emotional contagion to integrate user relation contexts in sentiment classification.

$$\begin{aligned} &\min_W \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ec_{ij}} \left\| \hat{Y}_{i*} - \hat{Y}_{j*} \right\|^2 \\ &= \min_W \sum_{k=1}^{c} \hat{Y}_{*k}^T (D_{ec} - A_{ec}) \hat{Y}_{*k} \\ &= \min_W tr(W^T X^T L_{ec} X W) \end{aligned} \tag{5}$$

where D_{ec} is a diagonal matrix, $D_{ec_{ii}} = \sum_{j=1}^{n} A_{ec_{ij}}$, $L_{ec} = D_{ec} - A_{ec}$.

Weak Dependency Connection Contexts

This paper uses communities to measure weak dependency connections between two non-directly connected microblogs. If two non-directly connected nodes are in the same community, a weak dependency connection exists between them.

Before modeling weak dependency connections, this paper first verifies whether the weak dependency connection theory holds true in the microblogging data. We first engage in a statistical study of the degree to which weak dependency connections and microblogs sentiment labels correlate, as a major motivation for our work is the intuition that microblogs connected by weak dependency connections tend to share the similar sentiment. We investigate the correlation between weak dependency connections and sentiment labels of microblogs from two types of statistical methods: the probability that two microblogs share the same sentiment conditioned on whether or not they are connected by weak dependency connections and a statistical hypothesis test on the correlation.

Figure 2 clearly shows that the probability of two microblogs connected by weak dependency connections sharing the same sentiment is much higher than chance on both HCR and OMD. It is noted that the positive correlation between weak dependencies and sentiment labels of OMD dataset is more stable than HCR dataset. This may be due to that there are more topics in the HCR dataset than in the OMD dataset and people usually have different views on different topics. The average number of microblogs related to a topic decreases as the number of topics increases, so using different community detection algorithms based on different mechanisms can lead to this phenomenon.

For the statistical hypothesis test, we form a null hypothesis: in terms of sentiment, there is no difference between relational data and random data. The sentiment difference between two microblogs is expressed by $T_{ij} = \|Y_{i*} - Y_{j*}\|_F^2$. In order to verify the validity of the weak relation theory, we establish two vectors wd_t and wd_r. Each element in wd_t represents the sentiment difference score between the i-th and the j-th microblog that are connected by a weak dependency connection. Elements in wd_r represent the sentiment difference score between two random selected microblogs. We perform a two-sample t test on the two vectors wd_t and wd_r. The null hypothesis is that there is no difference between the two vectors, i.e. $H_0 : wd_t = wd_r$. The alternative hypothesis is that the sentiment difference

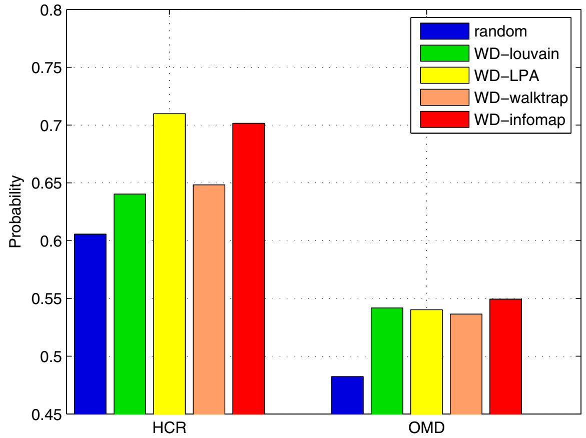


Figure 2: Shared sentiment probability conditioned on weak dependency connections

between microblogs with weak dependency connections is less than those without, $H_1 : wd_t < wd_r$. The t-test operations on both datasets reject the H_0 hypothesis with a significant level of 0.01. These two statistical results show that two microblogs connected by a weak dependency connection are more likely to have the same sentiment label. That is, the weak dependency connection theory holds true in microblogging data, which paves the way for our next study: how to exploit and model weak dependency connections into the microblog sentiment analysis system.

Next, we will explore how to introduce weak dependency connections into microblog sentiment analysis. In this paper, the matrix A_{wd} is used to represent the weak dependency connection matrix between microblogs. $A_{wd_{ij}} = 1$ if and only if the microblog m_i and m_j are in the same community. According to the theory that two microblogs in the same community are more likely to share the same sentiment, the Equation 6 model is established:

$$\begin{aligned}
&\min_W \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{wd_{ij}} \left\|\hat{Y}_{i*} - \hat{Y}_{j*}\right\|^2 \\
&= \min_W \sum_{k=1}^{c} \hat{Y}_{*k}^T (D_{wd} - A_{wd}) \hat{Y}_{*k} \\
&= \min_W tr(W^T X^T L_{wd} X W)
\end{aligned} \tag{6}$$

where D_{wd} is a diagonal matrix, $D_{wd_{ii}} = \sum_{j=1}^{n} A_{wd_{ij}}$, $L_{wd} = D_{wd} - A_{wd}$.

3.4. Incorporating Social Contexts

User contexts, user relation contexts, and weak dependency connection contexts are all social contexts. Through the t tests of user contexts, user relation contexts in Hu et al. [16] and the statistical results of weak dependency connections contexts in Section 3.3, we have reasons to believe the social contexts are likely helpful in microblog sentiment analysis. We use Equation 7 to combine the three contexts together.

$$\begin{aligned}
&\min_W \frac{\alpha_1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{sc_{ij}} \left\|\hat{Y}_{i*} - \hat{Y}_{j*}\right\|^2 + \frac{\alpha_2}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ec_{ij}} \left\|\hat{Y}_{i*} - \hat{Y}_{j*}\right\|^2 \\
&\qquad + \frac{\alpha_3}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} A_{wd_{ij}} \left\|\hat{Y}_{i*} - \hat{Y}_{j*}\right\|^2 \\
&= \min_W \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} (\alpha_1 A_{sc_{ij}} + \alpha_2 A_{ec_{ij}} + \alpha_3 A_{wd_{ij}}) \left\|\hat{Y}_{i*} - \hat{Y}_{j*}\right\|^2
\end{aligned} \tag{7}$$

where α_1, α_2 and α_3 are the model weights. In this paper, we set $\alpha_1 = 1$, $\alpha_2 = 1$ and $\alpha_3 = 1$. We set $A = \alpha_1 A_{sc} + \alpha_2 A_{ec} + \alpha_3 A_{wd}$, then Equation 7 can be transformed into Equation 8.

$$\begin{aligned} &\min_{W} \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} A \left\| \hat{Y}_{i*} - \hat{Y}_{j*} \right\|^2 \\ &= \min_{W} \sum_{k=1}^{c} \hat{Y}_{*k}^{T} (D - A) \hat{Y}_{*k} \\ &= \min_{W} tr(W^T X^T L X W) \end{aligned} \tag{8}$$

where D is a diagonal matrix, $D_{ii} = \sum_{j=1}^{n} A_{ij}$, $L = D - A$

3.5. Modeling Microblog Content

The popular method Least Squares is applied to fit the classification model for text information. In terms of multiclass classification tasks, the Least Squares aims to learn c classifiers by solving the optimization problem in Equation 9:

$$\min_{W} \frac{1}{2} \left\| XW - Y \right\|_F^2 \tag{9}$$

To represent texts for sentiment analysis better, Pang et al. [60] conducted experiments on different features to investigate their effectiveness. In their work, the unigram model with term presence as feature weight gets the best results. In addition, no stemming or stop-word lists are used in their work as some of them may carry sentiment information. Therefore, we do not perform stemming or remove stop-words and also use the unigram model. It is noted that our model is not confined to the unigram model. We can also use other text representation methods such as tfidf, word embedding and so on for specific sentiment classification tasks.

3.6. The Final Model

Unlike traditional text information, microblogs are short and have many noises which lead to a sparse matrix of unigrams. To handle this problem, we use sparse regularization L_1 norm to seek a sparse reconstruction of the feature space. To minimize the L_1 norm based linear reconstruction error can implement feature selection automatically and get a sparse representation of texts [62]. Thus, we add L_1 norm in our final model to get a more robust model. By combining the text content, the social contexts and L_1 norm, we get the final model (see Equation 10).

$$\min_W f(W;X,Y) = \min_W \frac{1}{2}\|XW - Y\|_F^2 + \frac{\alpha}{2} tr(W^T X^T LXW) + \beta \|W\|_1 \quad (10)$$

where α is the weight of social contexts in the model, β is the weight of regularization.

3.7. Learning

In this paper, we directly use the optimization algorithm in Hu et al. [16]. Motivated by [58], we propose to solve the non-smooth optimization problem in Equation 10 by optimizing its equivalent smooth convex reformulations. Firstly, Equation 10 can be reformulated by Equation 11 as a constrained smooth convex optimization problem.

$$\begin{aligned} &\min_{W\in Z} L(W;X,Y) = \frac{1}{2}\|XW - Y\|_F^2 + \frac{\alpha}{2} tr(W^T X^T LXW), \\ &where \quad Z = \{W \mid \|W\|_1 \le z\} \end{aligned} \quad (11)$$

$L(W;X,Y)$ is the differentiable part and Z is the non-differentiable part. $z \geq 0$ is the radius of the L_1-ball, and there is a one-to-one correspondence between β and z.

The smooth part of the optimization problem can be reformulated equivalently as a proximal regularization [59] of the linearized function $L(W;X,Y)$ at W_t, which is formally defined as:

$$\begin{aligned} &W_{t+1} = \arg\min_{W} G_{\lambda_t,W_t}(W) \\ &where \quad G_{\lambda_t,W_t}(W) = L(W_t;X,Y)+ \\ &< \bigtriangledown L(W_t;X,Y), W - W_t > + \frac{\lambda_t}{2}\|W - W_t\|_F^2 \end{aligned} \tag{12}$$

where λ_t is the step size in the t-th iteration. In this paper, the gradient of $L(W;X,Y)$ with respect to W can be computed using Equation 13.

$$\bigtriangledown L(W;X,Y) = X^T(XW - Y) + \alpha X^T LXW \tag{13}$$

When considering the constraints Z in Equation 11, and given β, the $(t+1)$-th W can be computed by Equation 14.

$$(W_{t+1})_{j*} = \begin{cases} (1 - \dfrac{\beta}{\lambda_t \|(U_t)_{j*}\|})(U_t)_{j*}, if \ \|(U_t)_{j*}\| \geqslant \dfrac{\beta}{\lambda_t} \\ 0, otherwise \end{cases} \tag{14}$$

where $U_t = W_t - \frac{1}{\lambda_t} \bigtriangledown L(W_t;X,Y)$.

4. Experiment

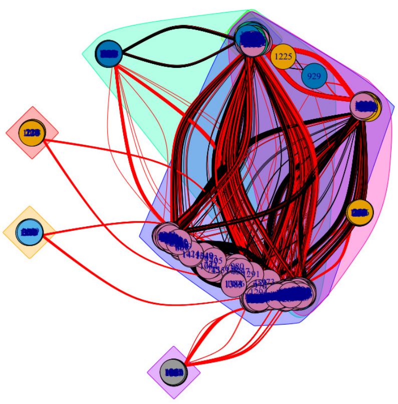
In order to verify the validity of the proposed model, we experiment with two datasets: HCR and OMD. We use accuracy, which is the proportion of true results (both true positives and true negatives) among the total number of cases examined, as a metric to measure the performance of different algorithms. The experiments are around: The effect of different contexts on sentiment analysis, the accuracy of various sentiment analysis methods and the influence of parameters on the accuracy.

4.1. Community Detection

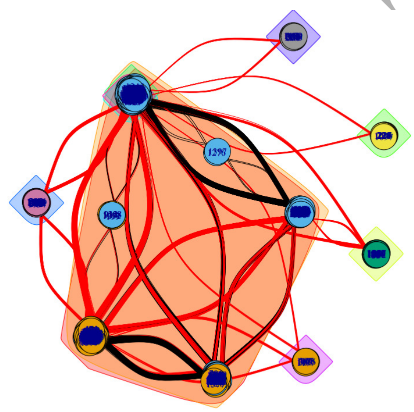
According to the model proposed by Section 3, we use louvain, LPA, walktrap and infomap to detect the communities and get weak dependency connections between microblogs. The results of the four algorithms on the HCR dataset and the OMD dataset are shown in Figure 3 and Figure 4 respectively. Nodes in the same community have the same background color.

4.2. Usefulness of Social Contexts

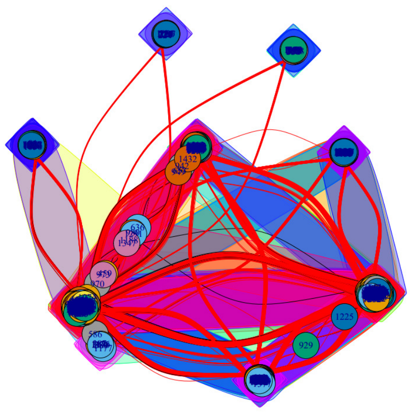
In this section, we compare the effects of different social contexts on the results of microblog sentiment analysis. We use 5-fold cross-validation in this part. None indicates that only text information is used for sentiment classification. SC represents the method using user contexts besides texts, EC represents that we add user direct relations into the text-based classification model. WD-XX indicates we add the weak dependency connection contexts into the text-based classification model. XX represents the algorithm used to explore the weak dependency connections between microblogs. All methods



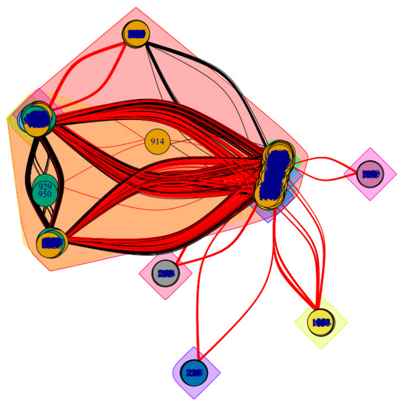
(a) Community Detection Result of Louvain



(b) Community Detection Result of LPA

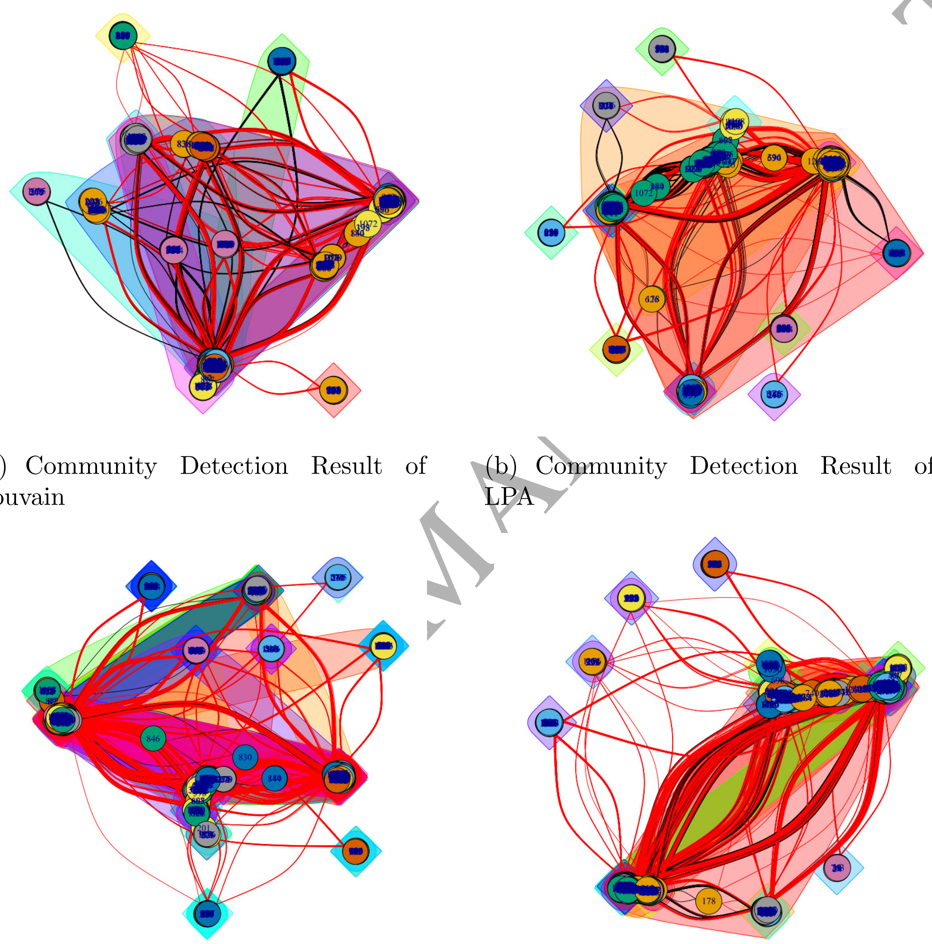


(c) Community Detection Result of Infomap



(d) Community Detection Result of Walktrap

Figure 3: HCR community detection result



(a) Community Detection Result of Louvain

(b) Community Detection Result of LPA

(c) Community Detection Result of Infomap

(d) Community Detection Result of Walktrap

Figure 4: OMD community detection result

are performed on both datasets. The results are shown in Figure 5. From the figure we can get the following observations:

1. Using social contexts can improve the accuracy of sentiment analysis on both datasets. The accuracy of the methods using social contexts is higher than the accuracy of using only text, which validates the usefulness of user contexts, user relation contexts and weak dependency connection contexts. Homophily does hold true in networks, that is the reason why using different social contexts can improve the performance of sentiment analysis. And in turn, this result can also be an experimental basis for the homophily theory.
2. User contexts get the lower improvement than the other social contexts. This is mainly due to A_{sc} is much sparse than other matrixes (A_{wd} and A_{ec}). For example, according to Table 2, each user in HCR dataset only has 1.78 tweets on average, while the average number of friends is 14.95.
3. Regardless of the HCR dataset or the OMD dataset, methods using weak dependency connection contexts get better performance than the other social contexts. Among them, the louvain-based algorithm gets the best performance, and the infomap-based method gets the lowest improvement on the accuracy of microblog sentiment analysis. The results indicate that weak dependency connection contexts are helpful for sentiment analysis. Weak dependency connection contexts can fully utilize the homophily theory of networks and find an effective way to represent the similarities between nodes, which is the reason behind its better performance than others.

To further explore why different community detection algorithms have different performances, we get the statistical information of the community detection results (shown in Table 3). Because we have no ground truth of community detection results on these datasets and it is not realistic to find a reasonable criterion to label such large-scale real network datasets by manual, so we use modularity [63] as a metric to measure the performance of community detection algorithm. As shown in Table 3, for both OMD and HCR, louvain has the smallest number of communities and the largest modularity, which may be the reason for its highest performance. Combining Figure 5 and Table 3, we find that both the number of communities and the modularity can impact the classification accuracy. The number of communities affects the sparsity of weak dependency matrix, while modularity affects the compactness within communities and the separation between communities. Larger modularity means that networks have denser connections between nodes within modules but sparser connections between nodes in different modules. Intuitively, a larger modularity and a relatively smaller number of communities can lead to a better performance, which is consistent with the results in this section.

Table 3: Statistics of community detection results

Algorithm		louvain	LPA	walktrap	infomap
Community Number	HCR	9	11	24	42
	OMD	13	22	68	71
Modularity	HCR	0.3275919	0.2417489	0.2709317	0.2610129
	OMD	0.4197237	0.3345575	0.3428504	0.3916175

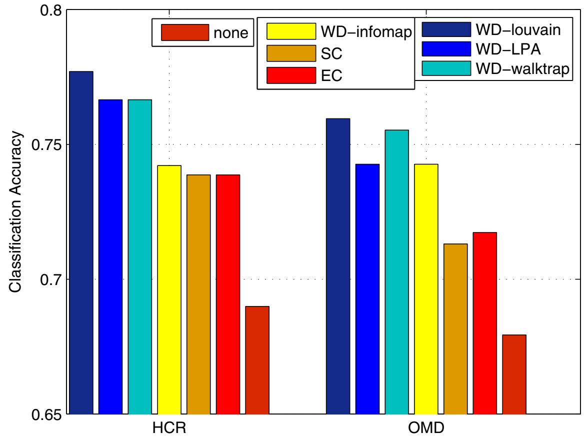


Figure 5: Microblog Sentiment Analysis accuracy with Different Contexts

4.3. Performance Evaluation

In this section, we use the random sampling method to test the accuracy of different methods on different sizes of training sets. The methods we use in this paper are listed below.

Least Squares (LS): Least Square method [9] is a widely used supervised classifier.

Lasso: Lasso [9] only use texts to identify sentiment. Comparing with Least Square method, Lasso adds $\|W\|_1$ to handle the sparse problem of classifier W, which we mentioned in Section 3.6.

MaxEnt: The Max Entropy classifier [12] is a probabilistic classifier which belongs to the class of exponential models. It is a widely used method for text classification.

Support Vector Machine (SVM): SVM [12] is a widely used classifier in the fields of text and hypertext categorization, images classification and so on.

Naive Bayes (NB): Like SVM, NB [12] is also a supervised classifier used in many fields.

SANT: A method proposed by [16] which combines sentiment consistency and emotional contagion.

SMSC: A method proposed by [40] which use graph information at the prediction stage.

M-louvain, M-LPA, M-walktrap and M-infomap are our proposed methods with different algorithms to get weak dependency connections.

In our method, there are two import parameters: α, β. The two parameters are all positive. α is the parameter that controls the contribution of social context information, β is the sparse regularization parameter. In this section, we set $\alpha = 0.0005$, $\beta = 1$ which are tuned by cross-validation. The training set and the test set are selected randomly from the original dataset to test our method. $p\%$ represents the percentage of the training set, and the rest $1-p\%$ are used for testing. Experimental results of HCR and OMD are shown in Figure 6 and Figure 7 respectively. In this part, we treat LS as the baseline.

Via comparing the results of different methods, we can draw the following observations.

1. Methods using social contexts such as SANT, SMSC and our proposed method have better performance than methods only using texts such as NB, MaxEnt, LS, Lasso and SVM in both HCR and OMD. We conducted two-sample one-tail t tests to compare the results of methods using social contexts with the results of methods only using texts. The results show

that methods using social contexts can significantly improve the classification accuracy with a significance level 0.01. NB, MaxEnt, LS, Lasso and SVM omit social contexts and they cannot handle the irony and sarcasm problems, so they behave worse than methods exploiting social contexts which solve irony and sarcasm to some extent. This phenomenon illustrates the validity of social contexts in sentiment analysis.

2. Our proposed method outperforms SANT and SMSC which also using social contexts on both datasets with different sizes of training data consistently. We also carried out t tests to compare the performance of different methods using social contexts. The experiment results demonstrate that our proposed model is able to achieve significant improvement (with the significance level 0.01) as compared to SANT and SMSC. SANT and SMSC only use user contexts and user relation contexts. In contrast, our method using weak dependency connections can deeply explore the relations between microblogs and the structure of microblog networks through the homophily theory. Therefore, our method can achieve better performance.
3. Lasso achieves better performance than LS, this implies using a sparse solution is an effective way to handle noisy microblog texts.
4. When there is only 50% data for training, our method outperforms the LS baseline with an 11.2820513% and 11.96754560% improvement on OMD and HCR respectively, better than that of both SANT and SMSC. This demonstrates that our method can improve the performance significantly when the number of labeled microblogs is small, which means we can save a lot of cost in labeling.

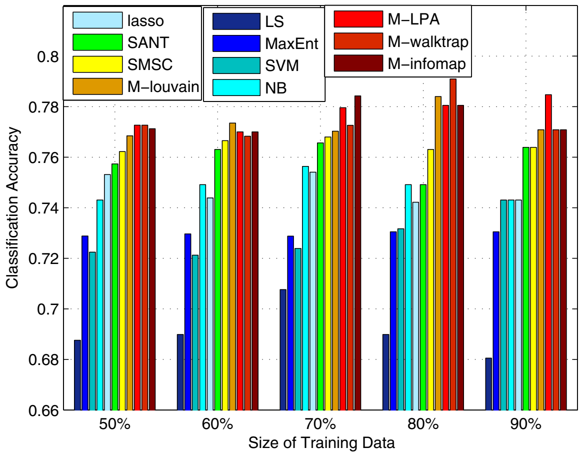


Figure 6: HCR sentiment accuracy

5. Methods using different community detection algorithms behave differently on different training sizes. Different from Section 4.2, we add user contexts and user relation contexts in the final model, which may weaken the effect of weak dependency connections. It is also noted that different community detection algorithms behave differently on different datasets. Therefore, to get better performance, it is essential to select an appropriate community detection algorithm for different datasets.

4.4. Parameter Analysis

In this subsection, we evaluate the effects of parameter selection of α and β on our methods. We use 80% of data on both datasets which are randomly selected for training, and the left data are used for testing. Figure 8 and 10 show the effect of α in detail when $\beta = 1$ on OMD and HCR respectively. Obviously, the performance of our proposed methods is not sensitive to the variation of α when the value of α is between 0.0001 and 0.0009. When α

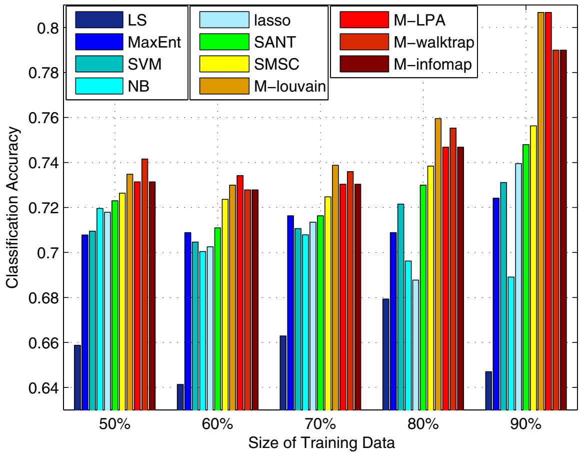


Figure 7: OMD sentiment accuracy

is too small, social contexts are not fully used in sentiment analysis. Thus, the performance increases as α increases from 0. However, when α is too large, the performance of the model mainly depends on social contexts, so it becomes worse.

Figure 9 and 11 show the performance of the proposed methods with the variation of β when $\alpha = 0.0005$. From Figure 9 and 11, it is noted that the performance of our model is not sensitive to β when β is in the range of $[0.1, 1]$. When $\beta \geq 1$, the performance of our proposed methods goes down as β gets bigger, but not too sharply. The reason behind the results is that when β is too large, the model relies on the sparse regularization too much and many features are filtered by the regularization. When β is too small, the sparse regularization is not fully used and many noises are remained in the training set. Therefore, in this situation, the performance of the proposed methods is still not very satisfactory.

It is an appealing property that our proposed model is not very sensitive to the variation of parameters as it can save a lot of time to tune parameters. Our method can consistently achieve good performance with a large range of parameter settings on both α and β.

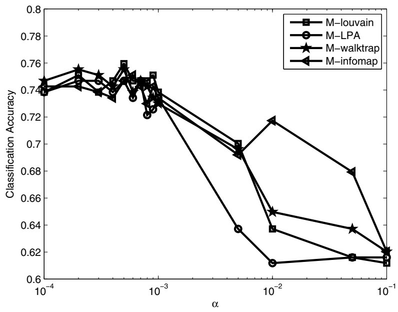


Figure 8: Sensitivity to Hyper-parameter α on OMD dataset

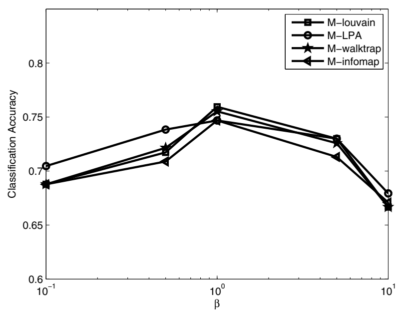


Figure 9: Sensitivity to Hyper-parameter β on OMD dataset

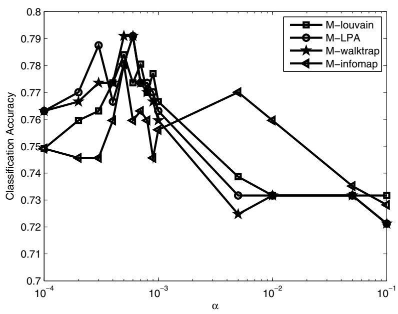


Figure 10: Sensitivity to Hyper-parameter α on HCR dataset

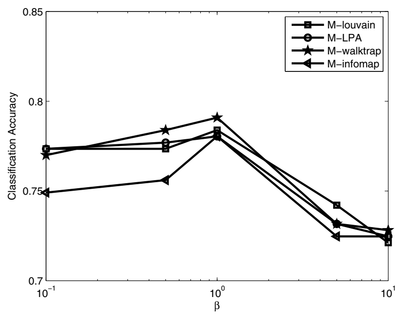


Figure 11: Sensitivity to Hyper-parameter β on HCR dataset

5. Conclusion and Discussion

Connections in online social networks are intrinsically heterogeneous and diverse relations are merged together. Nodes are likely to interact with similar nodes, which leads to forming communities in online social networks. Nodes in the same communities may not connect directly but there exist weak dependency connections between them. In this paper, we study how to exploit weak dependency connections for microblog sentiment analysis. We first adopt community detection methods to capture weak dependency connections, and then we build a new model which combines sentiment consistency, emotional contagion, and weak dependency connections together with text information. Experimental results on real-world datasets show that the proposed microblog sentiment analysis model outperforms the state-of-the-art models.

In this paper, we use Least Squares to model text information of microblogs. In future, we also want to extend Laplacian regularization to support vector machine (SVM) and maximum entropy model to see the differences between them. Deep learning methods have obtained very good performance across many different natural language processing tasks recently, so we are going to study how to combine social contexts with deep learning models in the near future.

Acknowledgement

This paper is supported by (1) the National Natural Science Foundation of China under Grant nos. 61672179, 61370083 and 61402126, (2) Research Fund for the Doctoral Program of Higher Education of China under Grant nos. 20122304110012, (3) the Youth Science Foundation of Heilongjiang Province of China under Grant no. QC2016083, (4) Heilongjiang postdoctoral Fund no. LBH-Z14071. This paper is also supported by China Scholarship Council.

References

[1] Abelson, R. P., 1983. Whatever became of consistency theory? Personality and Social Psychology Bulletin 9 (1), 37–54.

[2] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P10008.

[3] Bollen, J., Mao, H., Zeng, X., 2011. Twitter mood predicts the stock market. Journal of Computational Science 2 (1), 1–8.

[4] Cheng, K., Li, J., Tang, J., Liu, H., 2017. Unsupervised sentiment analysis with signed social networks. In: Proceedings of the thirty-first AAAI Conference on Artificial Intelligence and the twenty-ninth Innovative Applications of Artificial Intelligence conference. AAAI '17. AAAI Press, pp. 3429–3435.

[5] Clauset, A., 2005. Finding local community structure in networks. Physical review E 72 (2), 026132.

[6] Cui, A., Zhang, M., Liu, Y., Ma, S., 2011. Emotion tokens: bridging the gap among multilingual twitter sentiment analysis. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 238–249.

[7] Fersini, E., Pozzi, F. A., Messina, E., 2016. Approval network: a novel approach for sentiment analysis in social networks. World Wide Web, 1–24.

[8] Fiedler, M., 1973. Algebraic connectivity of graphs. Czechoslovak Mathematical Journal 23 (2), 298–305.

[9] Friedman, J., Hastie, T., Tibshirani, R., 2001. The elements of statistical learning. Vol. 1.

[10] Fu, M.-H., Chen, L.-Y., Lee, K.-R., Kuo, Y.-H., 2012. A novel opinion analysis scheme using social relationships on microblog. Springer Netherlands, Dordrecht, pp. 687–695.

[11] Girvan, M., Newman, M., JUN 11 2002. Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America 99 (12), 7821–7826.

[12] Go, A., Bhayani, R., Huang, L., 2009. Twitter sentiment classification using distant supervision. Technical report, Stanford Digital Library Technologies Project 1 (12).

[13] Granovetter, M., 1973. The strength of weak ties. American Journal of Sociology 78 (6), 1360–1380.

[14] Granovetter, M., 1983. The strength of weak ties: a network theory revisited. Sociological Theory, 201–233.

[15] Hatfield, E., Cacioppo, J. T., Rapson, R. L., 1993. Emotional contagion. Current Directions in Psychological Science 2 (3), 96–100.

[16] Hu, X., Tang, L., Tang, J., Liu, H., 2013. Exploiting social relations for sentiment analysis in microblogging. In: Proceedings of the sixth ACM

international conference on Web Search and Data Mining. ACM, pp. 537–546.

[59] Kernighan, B. W., Lin, S., Feb 1970. An efficient heuristic procedure for partitioning graphs. The Bell System Technical Journal 49 (2), 291–307.

[18] Kiritchenko, S., Zhu, X., Mohammad, S. M., 2014. Sentiment analysis of short informal text. Journal of Artificial Intelligence Research 50, 723–762.

[19] Kwak, H., Lee, C., Park, H., Moon, S., 2010. What is twitter, a social network or a news media? In: Proceedings of the 19th international conference on World Wide Web. WWW '10. pp. 591–600.

[20] Leung, I. X. Y., Hui, P., Lio, P., Crowcroft, J., JUN 2009. Towards real-time community detection in large networks. Physical Review E 79 (6, 2).

[21] Levin, D., Cross, R., NOV 2004. The strength of weak ties you can trust: the mediating role of trust in effective knowledge transfer. Management Science 50 (11), 1477–1490, 62nd Annual Meeting of the Academy-of-Management, DENVER, CO, AUG 09-11, 2002.

[22] Liu, K.-L., Li, W.-J., Guo, M., 2012. Emoticon smoothed language models for twitter sentiment analysis. In: Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence. AAAI '12. AAAI Press, pp. 1678–1684.

[23] Lu, T. J., Feb 2015. Semi-supervised microblog sentiment analysis using

social relation and text similarity. In: 2015 international conference on Big Data and Smart Computing (BIGCOMP). pp. 194–201.

[24] Newman, M., JUN 2004. Fast algorithm for detecting community structure in networks. Physical Review E 69 (6, 2).

[25] Palla, G., Derenyi, I., Farkas, I., Vicsek, T., JUN 9 2005. Uncovering the overlapping community structure of complex networks in nature and society. Nature 435 (7043), 814–818.

[26] Palla, G., Farkas, I. J., Pollner, P., Derenyi, I., Vicsek, T., JUN 28 2007. Directed network modules. New Journal of Physics 9.

[27] Pandarachalil, R., Sendhilkumar, S., Mahalakshmi, G. S., APR 2015. Twitter sentiment analysis for large-scale data: an unsupervised approach. Cognitive Computation 7 (2, SI), 254–262.

[28] Pons, P., Latapy, M., 2005. Computing communities in large networks using random walks. In: Computer and Information Sciences - ISCIS 2005, Proceedings. Vol. 3733 of Lecture Notes in Computer Science. SPRINGER-VERLAG BERLIN, HEIDELBERGER PLATZ 3, D-14197 BERLIN, GERMANY, pp. 284–293.

[29] Raghavan, U. N., Albert, R., Kumara, S., Sep 2007. Near linear time algorithm to detect community structures in large-scale networks. Physical Review E 76, 036106.

[30] Rosvall, M., Bergstrom, C. T., MAY 1 2007. An information-theoretic framework for resolving community structure in complex networks. Pro-

ceedings of the National Academy of Sciences of the United States of America 104 (18), 7327–7331.

[31] Rosvall, M., Bergstrom, C. T., JAN 29 2008. Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences of the United States of America 105 (4), 1118–1123.

[32] Shamma, D. A., Kennedy, L., Churchill, E. F., 2009. Tweet the debates: understanding community annotation of uncollected sources. In: Proceedings of the first SIGMM Workshop on Social Media. WSM '09. pp. 3–10.

[33] Speriosu, M., Sudan, N., Upadhyay, S., Baldridge, J., 2011. Twitter polarity classification with label propagation over lexical links and the follower graph. In: Proceedings of the first Workshop on Unsupervised Learning in NLP. EMNLP '11. pp. 53–63.

[34] Tan, C., Lee, L., Tang, J., Jiang, L., Zhou, M., Li, P., 2011. User-level sentiment analysis incorporating social networks. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge Discovery and Data Mining. KDD '11. pp. 1397–1405.

[35] Tang, J., Chang, Y., Liu, H., Jun. 2014. Mining social media with social theories: a survey. SIGKDD Explorations Newsletter 15 (2), 20–29.

[36] Tang, J., Wang, S., Hu, X., Yin, D., Bi, Y., Chang, Y., Liu, H., 2016. Recommendation with social dimensions. In: Proceedings of the thirti-

eth AAAI conference on Artificial Intelligence. AAAI '16. AAAI Press, pp. 251–257.

[37] Tang, L., Liu, H., 2009. Relational learning via latent social dimensions. In: Proceedings of the 15th ACM SIGKDD international conference on Knowledge Discovery and Data Mining. KDD '09. ACM, New York, NY, USA, pp. 817–826.

[38] West, R., Paskov, H. S., Leskovec, J., Potts, C., 2014. Exploiting social network structure for person-to-person sentiment analysis. Transactions of the Association for Computational Linguistics 2, 297–310.

[39] Wolfe, A., FEB 1997. Social network analysis: methods and applications . American Ethnologist 24 (1), 219–220.

[40] Wu, F., Huang, Y., Song, Y., JAN 29 2016. Structured microblog sentiment classification via social context regularization. Neurocomputing 175 (A), 599–609.

[41] Yang, D., Zhang, D., Yu, Z., Wang, Z., 2013. A sentiment-enhanced personalized location recommendation system. In: Proceedings of the 24th ACM conference on Hypertext and Social Media. HT '13. ACM, New York, NY, USA, pp. 119–128.

[42] Cambria, E., Mar 2016. Affective computing and sentiment analysis. IEEE Intelligent Systems 31 (2), 102–107.

[43] Cambria, E., Schuller, B., Xia, Y., White, B., Sep. 2016. New avenues in knowledge bases for natural language processing. Know.-Based Syst.

108 (C), 1–4.
URL http://dx.doi.org/10.1016/j.knosys.2016.07.025

[44] Severyn, A., Moschitti, A., 2015. Twitter sentiment analysis with deep convolutional neural networks. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, pp. 959–962.

[45] Ren, Y., Zhang, Y., Zhang, M., Ji, D., 2016. Context-sensitive twitter sentiment classification using neural network. In: AAAI. pp. 215–221.

[46] Wang, Y., Huang, M., Zhu, X., Zhao, L., 2016. Attention-based lstm for aspect-level sentiment classification. In: EMNLP. pp. 606–615.

[47] Muhammad, A., Wiratunga, N., Lothian, R., 2016. Contextual sentiment analysis for social media genres. Knowledge-based systems 108, 92–101.

[48] Cambria, E., Poria, S., Bajpai, R., Schuller, B. W., 2016. Senticnet 4: A semantic resource for sentiment analysis based on conceptual primitives. In: COLING. pp. 2666–2677.

[49] Poria, S., Peng, H., Hussain, A., Howard, N., Cambria, E., 2017. Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis. Neurocomputing 261, 217 – 230, advances in Extreme Learning Machines (ELM 2015).
URL http://www.sciencedirect.com/science/article/pii/S0925231217302023

[50] Bollen, J., Gonçalves, B., Ruan, G., Mao, H., 2011. Happiness is assortative in online social networks. Artificial life 17 (3), 237–251.

[51] Poria, S., Chaturvedi, I., Cambria, E., Hussain, A., 2016. Convolutional mkl based multimodal emotion recognition and sentiment analysis. In: Data Mining (ICDM), 2016 IEEE 16th International Conference on. IEEE, pp. 439–448.

[52] Fortunato, S., 2010. Community detection in graphs. Physics reports 486 (3), 75–174.

[53] Krishnamurthy, B., Wang, J., 2000. On network-aware clustering of web clients. In: Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication. SIGCOMM '00. ACM, New York, NY, USA, pp. 97–110.
URL http://doi.acm.org/10.1145/347059.347412

[54] Reddy, P. K., Kitsuregawa, M., Sreekanth, P., Rao, S. S., 2002. A graph based approach to extract a neighborhood customer community for collaborative filtering. In: International Workshop on Databases in Networked Information Systems. Springer, pp. 188–200.

[55] Wu, A. Y., Garland, M., Han, J., 2004. Mining scale-free networks using geodesic clustering. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 719–724.

[56] Cai, D., He, X., Han, J., Huang, T. S., 2011. Graph regularized nonneg-

ative matrix factorization for data representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (8), 1548–1560.

[57] Chung, F. R., 1997. Spectral graph theory. No. 92. American Mathematical Soc.

[58] Liu, J., Ji, S., Ye, J., 2009. Multi-task feature learning via efficient l 2, 1-norm minimization. In: Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence. AUAI Press, pp. 339–348.

[59] Kernighan, B. W., Lin, S., 1970. An efficient heuristic procedure for partitioning graphs. The Bell system technical journal 49 (2), 291–307.

[60] Pang, B., Lee, L., Vaithyanathan, S., 2002. Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, pp. 79–86.

[61] Hu, X., Sun, N., Zhang, C., Chua, T.-S., 2009. Exploiting internal and external semantics for the clustering of short texts using world knowledge. In: Proceedings of the 18th ACM conference on Information and knowledge management. ACM, pp. 919–928.

[62] Simon, N., Friedman, J., Hastie, T., Tibshirani, R., 2013. A sparse-group lasso. Journal of Computational and Graphical Statistics 22 (2), 231–245.

[63] Newman, M. E., 2006. Modularity and community structure in networks. Proceedings of the national academy of sciences 103 (23), 8577–8582.