Adding the Sentiment Attribute of Nodes to Improve Link Prediction in Social Network

Shaoliang Shi
School of Computer Science and Engineering
Guilin University of Electronic Technology
Guilin, China
E-mail:shishaoliang123@163.com

Yimin Wen*
Guangxi Key Laboratory of Trusted Software
Guilin University of Electronic Technology
Guilin, China
E-mail: ymwen2004@aliyun.com

Yunpeng Li
College of Business Administration
Capital University of Economics and Business
Beijing, China
E-mail: dr_lyp@163.com

Wu Xie
School of Computer Science and Engineering
Guilin University of Electronic Technology
Guilin, China
E-mail: bailange587@126.com

***Abstract*—Link prediction is an important tool for many social media sites to find the missing and future links among users. Understanding users' sentiment and their social relationships are potentially valuable. In this paper, two new sentiment similarity measures have been proposed and an algorithm has been designed to do link prediction by incorporating the structure and sentiment attribute of nodes. In order to evaluate the proposed algorithm, links and tweets with regard to some hot topics of 2014 FIFA World Cup Brazil are crawled from Tencent Weibo, and the sentiment distributions of crowds are analyzed for each topic. The experimental results show that the number of users with the same emotion and the sentiment distributions of crowds will influence a user to link with another user, so the sentiment attribute of nodes in social network can help to improve the performance of link prediction.**

Keywords-link prediction, sentiment analysis, social networks, micro-blogging, sentiment distribution.

I. INTRODUCTION

Social media site has become one kind of popular online communication tool that improves the possibility of communication between users. With the popularity of micro-blogging sites on the Internet, such as Twitter[1] and Tencent Weibo[2], millions of people all over the world have become their users. People can freely share information, opinions, knowledge, insights, emotion and experiences by taking advantage of the open and effective nature of these micro-blogging systems. The main activity of micro-blogging is tweet that a user post a short message[1]. And further, users can form a clear social network by following another user and thus automatically receive his/her tweets. These facts implies the rapid growth and changes over time in underlying structure (nodes and links) of these social networks [2], [14].

In recent years, the study of link prediction in social network has drawn increasing attention from the data mining community. Link prediction can predict the missing and future links between two nodes, which has important theoretical and practical value [3]. In the aspect of theory, researches on link prediction can help us understand the evolution mechanism of complex network. In practical application, link prediction can be used in social network to recommend friends for improving user experience as well as reducing the risk of information overload.

In order to improve link prediction, many algorithms have been proposed. The simplest method of link prediction is the similarity-based algorithm, there are three sub-types of the similarity-based algorithm. The first one is based on the local structure of a network, such as Common Neighbors[4], Jaccard Coefficient[4], Preferential Attachment[5], Adamic Adar[6], and Resource Allocation[7]. The second one is based on the overall structure of a network [2],[3]. Additionally, some studies took into account the semantic similarity of nodes to improve the performance of prediction.

However, these studies just considered the network structure or only focused on the semantic information of user generated content while ignored sentiment information. In social sciences, it is well-established that emotion or sentiment play a distinct role in social life [8]. Sha[9] has pointed out that emotion plays a key role in the interpersonal communication and emphasized that the role of emotion is more reliable and stronger than information communication between people.

Recently, Thelwall[10] found the existence of emotion homophily among social network. Tan et al.[11] utilized the information of user-user relationships to improve user-level sentiment analysis in online social networks, which showed that the probability of two connected users sharing the same sentiment on a topic is much higher than random users. Hu et al.[8] proposed a model for document-level sentiment classification, which showed that the user-centric social relations are helpful for sentiment classification. In these

* To whom correspondence should be addressed.
[1] https://twitter.com/
[2] http://t.qq.com/

978-1-4673-7682-2/15/$31.00 ©2015 IEEE

studies, the principle of homophily was used to improve the effect of sentiment analysis for different level. However, on the contrary, is it much more likely for two users to be connected if they share similar sentiment ?

In order to tackle the above issue, we propose an algorithm that integrate the local structure information of a nxetwork and sentiment attribute of nodes to conduct link prediction over a Tencent Weibo dataset including ten hot topics. To begin with, we constructed a graph with the networks data and calculate several similarity metrics using the local information from the network structure. Then, an emotion extraction model is built and two sentiment measures are defined for sentiment similarity computation. Finally, we consider the link prediction as a binary classification problem by training decision tree model to predict the potential link between users. In addition, we further analyze the relationship between the distribution of sentiment score and the performance of link prediction.

The rest of the paper is organized as follows. Section II states the problem of link prediction and introduces the structure measures used to do link prediction. In Section III, we discuss the related work on link prediction. The algorithm of sentiment extraction and the proposed algorithm to do link prediction are presented in Section IV. Then Section V describes the experiment and results. Finally, we finish the paper with conclusions and plans for future work.

II. LINK PREDICTION

A. Problem Statement

A social network can be modeled as a graph $G(V, E)$, where V is the set of nodes and E is the set of the edges between nodes. The complete graph, denoted by U, containing all $|V|(|V|-1)/2$ possible links among all nodes, where $|V|$ denotes the number of elements in V. Then, the set of nonexistent link is U-E. The task of link prediction is to find out whether a nonexistent link will become true or not [4],[14].

In this paper, link prediction problem is transformed into a binary classification problem: linking (positive class) and not-linking (negative class). Some performance metrics are defined for prediction evaluation. Like with the method used in [12], for each feature vector, a predictor p can make either a positive (P) or a negative (N) prediction concerning the corresponding label. In the positive case, if p is correct, the prediction is said to be true-positive (TP), otherwise it is false-positive (FP). Conversely, in the negative case a prediction can be either true-negative (TN) if correct or false-negative (FN) if wrong. Now, the metrics of precision and recall can be defined as in (1), (2) .

$$\text{Precision}=|TP|/(|TP|+|FP|) \quad (1)$$

$$\text{Recall}=|TP|/(|TP|+|FN|) \quad (2)$$

Based on them, another metric called the F-Measure is introduced to numerically compare predictors. It is the harmonic mean of the precision and recall.

$$\text{F-Measure}=2*\text{Precision}*\text{Recall}/(\text{Precision}+\text{Recall}) \quad (3)$$

B. Structural Measures

The basic structural definition for a node $x \in V$ in the network is its neighborhood $\Gamma(x) = \{y \mid (x, y) \in E \vee (y, x) \in E\}$. Based on this definition, several structural measures are introduced [2], [4], [14].

1) Common Neighbors(CN): In common sense, two nodes, x and y, are more likely to have a link if they have many common neighbors. So, the common neighbors measure refers to the size of the set of all common friends between x and y, according to Equation 4.

$$S_{xy}^{CN} = |\Lambda_{xy}| = |\Gamma(x) \cap \Gamma(y)| \quad (4)$$

where $|Q|$ is the cardinality of the set Q, and Λ_{xy} denotes the set of common neighbors.

2) Jaccard Coefficient(JAC): This measure indicates whether two nodes of a network have a significant number of common neighbors regarding their total neighbors set size. It was proposed by Jaccard over a hundred years ago, for an undirected network it is defined according to Equation 5.

$$S_{xy}^{JAC} = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|} \quad (5)$$

3) Preferential Attachment(PA): The mechanism of preferential attachment can be used to generate evolving scale-free networks, where the probability that a new link will connect x and y is proportional to $|\Gamma(x)| \times |\Gamma(y)|$. Motivated by this mechanism, the corresponding similarity index can be defined as Equation 6.

$$S_{xy}^{PA} = |\Gamma(x)| \times |\Gamma(y)| \quad (6)$$

4) Adamic Adar(AA): This measure refines the simple counting of common neighbors by assigning more weight to the less-connected neighbors, as defined in Equation 7.

$$S_{xy}^{AA} = \sum_{z \in \Lambda_{xy}} \frac{1}{\log |\Gamma(z)|} \quad (7)$$

5) Resource Allocation(RA): This index is motivated by the resource allocation dynamics on complex networks. Consider a pair of nodes, x and y, which are not directly connected. The node x can send some resource to y, with their common neighbors playing the role of transmitters. In the simplest case, It is assumed that each transmitter has a unit of resource, and will equally distribute it to all its neighbors. The similarity between x and y can be defined as the amount of resource y received from x.

$$S_{xy}^{RA} = \sum_{z \in \Lambda_{xy}} \frac{1}{|\Gamma(z)|} \quad (8)$$

III. RELATED WORK

To date, many link prediction algorithms have been proposed. It can be classified roughly into three categories: the similarity-based algorithm, the maximum likelihood method and the probabilistic model.

The similarity-based algorithm computes the similarity between a pair of nodes. There are three sub-types in this kind of algorithm. One type is based on the local structural information of a network. Liben-Nowell et al.[2] and Zhou et al. [7] systematically compared a number of local similarity indices in real networks, such as CN, JAC, AA, RA, PA. According to their experimental results, the RA index performs best, while AA and CN indices have the second best overall performance among all the above mentioned local indices. Another type is based on the overall structural information a network, Such as Katz Index, SimRank[4]. Although the global indices can provide much more accurate prediction than the local ones, their calculation are usually time consuming. In addition, with the rapid development of online social networks, the interactive content are valuable for the establishment of friend relationships[13]. So, some algorithms were proposed for semantic similarity analysis using user generated content [1], [12].

Maximum likelihood method presuppose some organizing principles of the network structure, with the detailed rules and specific parameters obtained by maximizing the likelihood of the observed structure. Then, the likelihood of any non-observed link can be calculated according to those rules and parameters. It mainly contains hierarchical structure model and stochastic block model[15].

Probabilistic model aims at building a model with a group of adjustable parameters using the attributes of nodes, and then optimize it to find the optimal parameters so as to make the model best fit the observed data of the network. There are three mainstream methods, respectively called Probabilistic Relational Model (PRM)[16], Probabilistic Entity Relationship Model (PERM)[17] and Stochastic Relational Model (SRM)[18]. In the ideal case, probabilistic model can get higher accuracy in link prediction, but its high computational complexity make the range of its application limited.

IV. LINK PREDICTION INCORPORATING THE STRUCTURE AND SENTIMENT

A. Sentiment Extraction

Sentiment analysis is one of the key emerging technologies in the effort to help people explore the huge amount of user generated content in social network[11]. Different from the texts in traditional media, micro-blogging texts are noisy, short, and informal, which bring a huge challenge to sentiment analysis. A large amount of work was dedicated to the problem of sentiment analysis from different perspective in recent years[19]. Therefore, the sentiment extraction models proposed in[20], [21] are adopted in this paper.

Emotional words are employed to compute the sentiment value of a sentence. So we collect a Chinese sentiment dictionary from DataTang[3] . This dictionary consists of 23419 emotional words and its corresponding sentiment value. Moreover, in order to much more accurately analyze the sentiment of a sentence, more factors are considered in the process of calculation. For example, a synonym list and a set of stop words are also obtained from DataTang. The negation word list and adverb list are constructed by manual.

At first, micro-blogging texts are segmented using the Chinese lexical analysis system ICTCLAS[4] . Then, the part-of-speech tags are used to extract the underlying emotional words. The adjective, verb, noun and adverb words may be emotional word in Chinese. We determine emotional word through finding whether it or its synonym is in the Chinese sentiment dictionary or not.

In Chinese, some words can reverse the meanings of sentence, we call it negation word, such as no, not, never, neither, impossible, etc. If they modify an emotional word, the emotional polarity of the emotional word will be reversed.

In addition, degree adverb (e.g., very) and exclamation marks (e.g., !) are also considered to impact on the sentiment value of an emotional word. So, if an emotional word is modified by these words, its sentiment value will be strengthened or weakened.

The algorithm of sentiment extraction is shown in Fig.1.

Algorithm 1: Sentiment Extraction

Input: T, the micro-blogging text sets consisting of all on-topic tweets of a user

Output: S_u, the sentiment score of a user

1. For each micro-blogging p in T, it is split using ICTCLAS, delete stop words and filter out non-emotional words according to the part-of-speech tags.
2. For each word in p, decide whether it's an emotional word.
3. For each negation word w_i in p, find the nearest emotional word, and reverse its sentiment value from $S_{w_{i+1}}$ to $-S_{w_{i+1}}$.
4. For each degree adverb in p, find the nearest emotional word, and multiply its sentiment value with the corresponding coefficient α.
5. Calculate the sentiment value of p by the function of emotional words.

$$S_p = \sum_{i=1}^{m} \alpha * S_{w_i}$$

where S_p , S_{w_i} means the sentiment value of micro-blogging p, and the sentiment value of emotional word w_i, respectively. And m is the total number of emotional words in the micro-blogging p.

6. If there exists an exclamation mark at the end of the microb-logging p, the sentiment value of p will multiply a corresponding coefficient β.

$$S_p = \beta * S_p$$

7. Then, the sentiment value of a user is defined as the average of all S_p.

$$S_u = \sum_{i=0}^{n} S_p / n$$

where n is the number of micro-blogging text in T.

8. Return S_u.

Figure 1. The algorithm of sentiment extraction

[3] http://www.datatang.com/

[4] http://ictclas.nlpir.org/

In this paper, the corresponding values of coefficient α are referred to the Table I and the coefficient β is set as 1.3. It is consistent with the literature as in[21].

TABLE I. THE VALUES OF α

Degree adverbs	Coefficient α
very, most, best, particularly, etc.	1.5
more, further, fairly, etc.	1.2
a little, a few, a bit, etc.	0.8
seldom, hardly, rarely, etc.	0.5

B. The Proposed Method

After getting the sentiment score of all users on a specific topic, we propose two measures to compute the sentiment similarity between two users(nodes).

Emotional Polarity (EP): In this paper, we calculate the similarity of a pair of nodes by utilizing the emotional polarity between two nodes. Here, we only consider that whether the polarity of these nodes is same or not, regardless of the sentiment value. The detailed definition of EP is in (9), S_x, S_y denotes the sentiment values of the node x and y, respectively.

$$S_{xy}^{EP} = \begin{cases} \text{pos}, & S_x>0 \quad \text{and} \quad S_y>0; \\ \text{neg}, & S_x<0 \quad \text{and} \quad S_y<0; \\ \text{neutral}, & S_x=0 \text{ and } S_y=0; \\ \text{diff}, & \text{otherwise.} \end{cases} \tag{9}$$

Emotional Similarity(ES): The emotional similarity is defined as "emotional distance", which is the absolute value of the difference between the sentiment values of a pair of nodes if their values have the same polarity, otherwise, the emotional similarity is specified as 100 which is an impossible sentiment value in experiment. That is, the smaller the emotional distance is, the more similar the pair of nodes will become. The specific definition is in (10). The sgn (S_x) is a symbol function of the S_x.

$$S_{xy}^{ES} = \begin{cases} |S_x - S_y|, & \text{sgn}(S_x)=\text{sgn}(S_y); \\ 100, & \text{otherwise.} \end{cases} \tag{10}$$

After obtained the measures based on the local structural information of a social network, and the emotional polarity as well as emotional similarity between each pair of nodes, we considered these measures as the feature values that describe the similarity between two nodes. Then, a feature vector can be constructed for each pair of nodes as {CN, JAC, PA, AA, RA, EP, *label*} or {CN, JAC, PA, AA, RA, ES, *label*}. If there is a link between a pair of nodes, the corresponding feature vector is labeled as *TRUE*, or not labeled as *FALSE*.

The proposed algorithm is showed in Fig. 2.

V. EXPERIMENT AND RESULTS

A. Data Collection

In this paper, the experiment datasets are crawled from the Tencent Weibo which launched at April 2010 in China [1]. We developed a crawler through the API of Tencent Weibo and crawled 32 hot topics in June and July, which are related to the team names who joined the 2014 World Cup Brazil.

Algorithm 2: Link prediction incorporating the structure and sentiment attribute

Input: social network G, the parameter c, and T the microblogging text sets consisting of all on-topic tweets of all users in G

Output: link prediction

1. According to the network graph G, calculate the local structural measures for each pair of nodes, including CN, JAC, PA, AA, and RA;
2. Call the algorithm of sentiment extraction to get the sentiment value of each node;
3. Compute the sentiment similarity, including EP, ES, for each pair of nodes;
4. Case c
   a) *c==EP*: the feature vectors for each pair of nodes are set as { CN, JAC, PA, AA, RA, ***EP***, *label*} ;
   b) *c==ES:* the feature vectors for each pair of nodes are set as { CN, JAC, PA, AA, RA, ***ES***, *label*}；
   END
5. Generating training and test dataset;
6. Training a classifier on training dataset;
7. Using the trained classifier to do link prediction.

Figure 2. The proposed algorithm of link prediction

The title of a topic is composed of some keywords which are enclosed by "#" in Tencent Weibo. At first, we crawled all the source file of the on-topic tweets using the API function *StatusesAPI.htTimelineExt* [5] in which the topic title and the authentication information are input as parameters. Then, we extracted the on-topic tweets and the username who wrote the tweets from the source file. In addition, we collected all of the reciprocal friend relationships for each user who pay attention to the topic using the function *FriendsAPI.mutualList* [6] . Finally, we constructed a graph using the users and their friendships we crawled.

To ensure the data is sufficiently reliable and available for experiments, several preprocessing steps were performed. At first, the singleton nodes are pruned, because these nodes didn't have any friend in the constructed networks. The number of the users with link in some topic are shown in Fig. 3. It can be seen that the number of users with link is less than 100 in most topics, we don't take these topics into account in experiments. In addition, we filtered out some invalid tweets, because it's difficult to extract the sentiment value of these tweets. For example, some tweets may just contain one or two words, URLs, and some information that a user refers to other user using "@". The overview of the experimental datasets is provided in Table II.

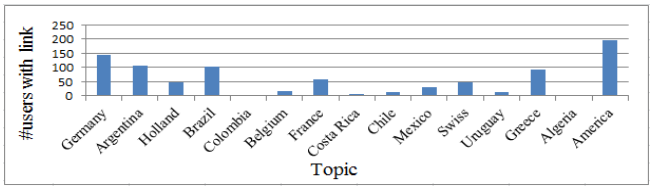

Figure 3. The number of users with link in each topic

[5] https://open.t.qq.com/api/statuses/ht_timeline_ext
[6] https://open.t.qq.com/api/friends/mutual_list

TABLE II. STATISTICS FOR EXPERIMENTAL DATASETS

Topic	#on-topic tweets	#users in topic	#users with link	#links in topic
Germany	2977	2053	144	712
Argentina	2446	1002	106	302
Brazil	1959	1014	104	198
USA	17821	4950	195	203
Italy	4113	810	110	178
Spain	4594	2107	175	241
Ghana	11898	11178	279	252
England	1232	632	102	302
Korea	30933	4440	182	140
Japan	22314	1754	119	145

B. *Dataset Generation*

Social network datasets are generally imbalanced, because the percentage of not-linking is huge. For example, the network corresponding to the topic Germany includes 144 nodes, and then there will be C_{144}^{2} possible links. However, there just exist 712 links in the real network. Hence, the strategy of under-sampling by deleting some negative examples is taken to rebalance the datasets. In the experiments, we made a random sampling according to the ratio of 2:1, which the not-linking sample is twice as large as linking sample in final dataset.

In order to avoid the unrepresentative of dataset, random sampling is repeated 100 times to get different datasets and the 10-fold cross validation is used to estimate the performance for each dataset in this paper.

C. *Experiment Setup*

In this paper, we use the J48 (the Weka[7] adaptation of C4.5 algorithm) for supervised learning. To test the effectiveness of the proposed approach, an algorithm is used as a baseline for comparing, in which the feature vectors are constructed as {CN, JAC, PA, AA, RA, *label*}[22].

For convenience, the baseline algorithm is denoted as 'Base', and the two case of the proposed algorithm are denoted as 'Base&EP' and 'Base&ES' respectively.

D. *Experimental Results*

Precision, Recall and F-Measure metrics are employed to validate the quality of each link prediction algorithm.

Looking from Table III, one can notice that the proposed algorithm outperforms the baseline method for each topic according to each metric. Even the promotion on each topic is small, all of these results illustrate that the algorithm which adding emotional factor can improve link prediction. Further more, it's worth noting that Base&ES method performs better than Base&EP method on each topic. So can we think that the sentiment attribute of nodes in social network can promote the performance of link prediction? i.e., it is much more likely for users to be connected if they share a similar emotion on the same topic than they differ.

Therefore, what kind of relationship between the emotional information and the formation of link?

[7] http://www.cs.waikato.ac.nz/ml/weka/

TABLE III. THE PERFORMANCE OF EACH LINK PREDICTION ALGORITHM

Topics	Methods	Precision	Recall	F-Measure
Germany	Base	0.9058±0.0000	0.9063±0.0000	0.9055±0.0000
	Base&EP	0.9062±0.0000	0.9067±0.0000	0.9062±0.0000
	Base&ES	0.9071±0.0000	0.9077±0.0000	0.9069±0.0000
Spain	Base	0.9025±0.0001	0.9017±0.0001	0.9020±0.0001
	Base&EP	0.9032±0.0001	0.9022±0.0001	0.9024±0.0001
	Base&ES	0.9043±0.0001	0.9029±0.0001	0.9035±0.0001
Korea	Base	0.9043±0.0001	0.9035±0.0001	0.9016±0.0001
	Base&EP	0.9057±0.0001	0.9043±0.0001	0.9021±0.0001
	Base&ES	0.9062±0.0001	0.9056±0.0001	0.9039±0.0001
England	Base	0.8878±0.0001	0.8865±0.0001	0.8929±0.0001
	Base&EP	0.8950±0.0001	0.8951±0.0001	0.8949±0.0001
	Base&ES	0.8975±0.0001	0.8950±0.0001	0.8958±0.0001
Japan	Base	0.8606±0.0002	0.8614±0.0002	0.8584±0.0002
	Base&EP	0.8625±0.0003	0.8634±0.0003	0.8604±0.0003
	Base&ES	0.8627±0.0002	0.8637±0.0002	0.8622±0.0002
Italy	Base	0.8693±0.0001	0.8657±0.0001	0.8600±0.0001
	Base&EP	0.8708±0.0001	0.8667±0.0001	0.8610±0.0002
	Base&ES	0.8799±0.0002	0.8656±0.0002	0.8642±0.0001
USA	Base	0.8805±0.0002	0.8809±0.0002	0.8803±0.0002
	Base&EP	0.8821±0.0001	0.8816±0.0001	0.8828±0.0001
	Base&ES	0.8840±0.0002	0.8835±0.0002	0.8858±0.0002
Ghana	Base	0.7529±0.0003	0.7582±0.0002	0.7507±0.0003
	Base&EP	0.7547±0.0002	0.7593±0.0002	0.7520±0.0003
	Base&ES	0.7589±0.0003	0.7642±0.0002	0.7586±0.0003
Argentina	Base	0.8391±0.0001	0.8409±0.0001	0.8375±0.0001
	Base&EP	0.8455±0.0001	0.8462±0.0001	0.8451±0.0001
	Base&ES	0.8456±0.0001	0.8470±0.0001	0.8455±0.0001
Brazil	Base	0.8922±0.0001	0.8916±0.0001	0.8901±0.0001
	Base&EP	0.8963±0.0001	0.8968±0.0001	0.8956±0.0001
	Base&ES	0.8983±0.0002	0.8991±0.0002	0.8990±0.0002

In order to further analyze the experimental results, Fig. 4 shows the distribution of sentiment score for every topic. From these figures, even the number of people in some topics are not large, some interesting phenomenon can be obviously seen that there are some peak points for each topic, the positive peak is always taller than the negative peak, and only a minority of people will express stronger emotion.

Table IV shows the relationship between DPerform(the difference of F-Measure between Base and Base&ES method) and DPeak(the difference of sentiment score between the positive and the negative tallest peak). It can be seen that under most topics there is an interesting rule—with the increasing of DPeak, the DPerform increases accordingly. So, we guess the reason why the emotional factor can improve link prediction lies in that the greater the number of people have the same emotion, the more possibility they will be friends, and the larger the difference between the positive emotion and the negative emotion which express by two crowds of people respectively, the more likely they will become friends.

TABLE IV. THE RELATIONSHIP BETWEEN DPERFORM AND DPEAK.

Topic	Germany	Spain	Korea	England	Japan
DPeak	1.9310	2.1610	2.2850	3.0070	3.0800
DPerform	0.0014	0.0015	0.0023	0.0029	0.0038
Topic	**Italy**	**USA**	**Ghana**	**Argentina**	**Brazil**
DPeak	3.5160	3.3050	3.5690	3.7770	3.8960
DPerform	0.0042	0.0055	0.0079	0.0080	0.0089

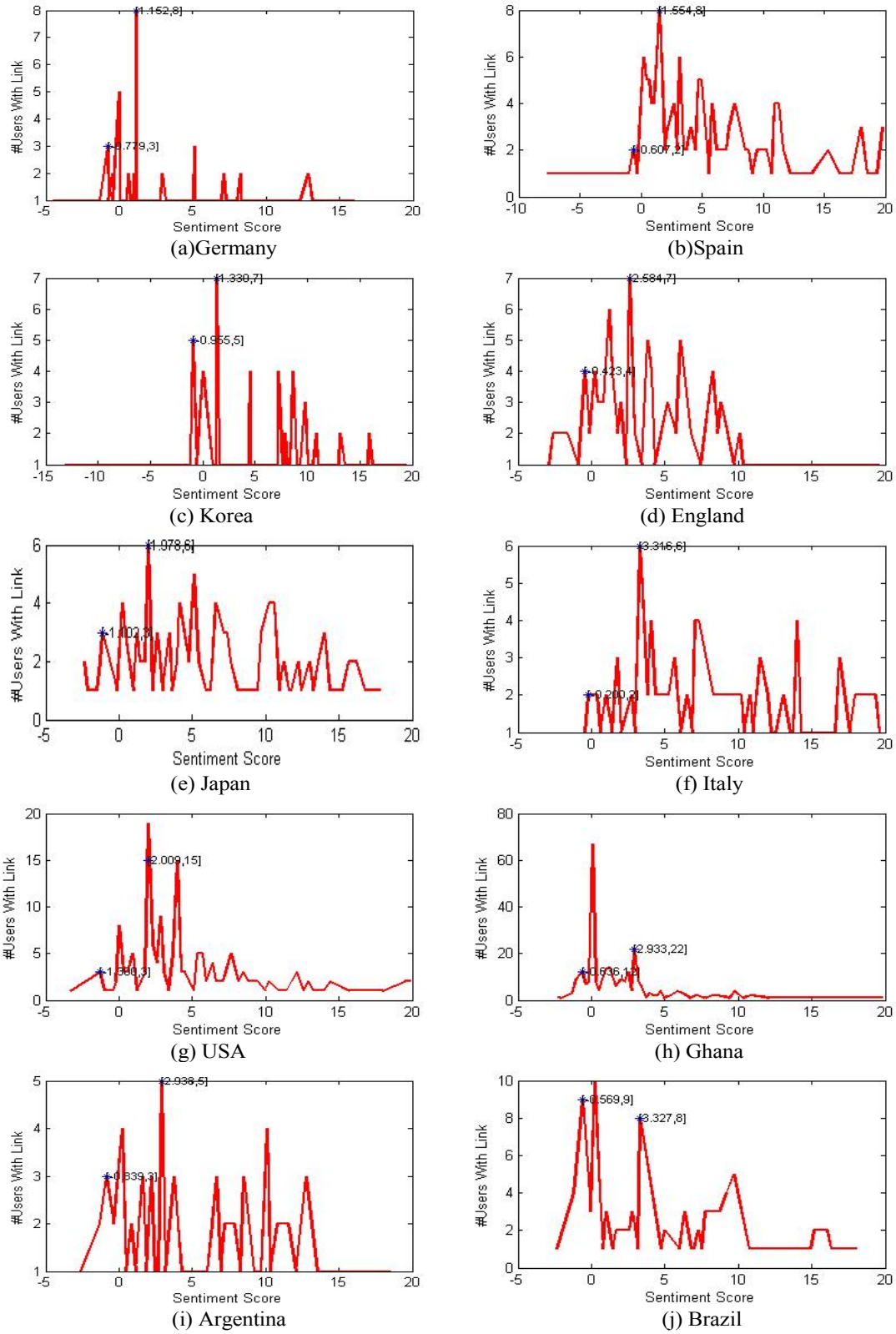

Figure 4. The distribution of sentiment score

VI. Conclusion and Future Work

Nowadays, many kinds of social media are more and more profoundly affects the life of people, so in this paper, an algorithm has been proposed to do link prediction incorporating the structure and sentiment attribute of nodes. In order to evaluate the proposed algorithm, links and tweets with respect to several hot topics are crawled and the sentiment distributions of several crowds are analyzed. The experimental results show that the proposed methods get better performance than the baseline method, i.e., the sentiment attribute of nodes in social network can help to improve the performance of link prediction.

Recently, a new interesting algorithm was proposed to use sentiment homophily to do link prediction through a factor graph model[23]. Different from this algorithm, our work assumed that the entire links are independent and identically distributed, and the emotions of people are considered only on one topic.

We intend to improve the proposed algorithm from three aspects in the near future. Firstly, more topics are needed to build more accurate prediction model. Secondly, we will try to take a consideration of predicting future link. Thirdly, we want to apply this model to the actual network for user recommendation.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (61363029,61163057,71340025), Guangxi Scientific Research and Technology Development Project (14124005-2-1), Innovation Project of GUET Graduate Education (GDYCSZ201468), and Guangxi Key Laboratory of Trusted Software (KX201311).

References

[1] H. Wu, V. Sorathia, and V. K. Prasanna, "Predict whom one will follow: followee recommendation in microblogs," In: 2012 International Conference on Social Informatics (SocialInformatics), Lausanne, pp. 260-264, 2012.

[2] D. Liben-Nowell, and J. Kleinberg, "The link-prediction problem for social networks," Journal of the American society for information science and technology. Vol. 58(7), pp. 1019-1031, 2007.

[3] Y. Yu, and X. Wang, "Link prediction in directed network and its application in microblog," Mathematical Problems in Engineering 2014.

[4] L. Lü, and T. Zhou, "Link prediction in complex networks: A survey," Physica A: Statistical Mechanics and its Applications, vol. 390(6), pp. 1150-1170, 2010.

[5] Xie Y B, Zhou T, Wang B H. Scale-free networks without growth[J]. Physica A: Statistical Mechanics and its Applications, 2008, 387(7): 1683-1688.

[6] Adamic L A, Adar E. Friends and neighbors on the web[J]. Social networks, 2003, 25(3): 211-230.

[7] T. Zhou, L. Lü, and Y. C. Zhang, "Predicting missing links via local information," Eur. Phys. J. B. vol. 71(4), pp. 623-630, 2009.

[8] X. Hu, L. Tang, J. L. Tang,and H. Liu, "Exploiting social relations for sentiment analysis in microblogging," In: Proceedings of the sixth ACM international conference on Web search and data mining, New York, pp. 537-546, 2013.

[9] L. X. Sha, Social Psychology. China Renmin University Press, Beijing, 1995.

[10] M. Thelwall, "Emotion homophily in social network site messages," First Monday, 2010.

[11] C. Tan, L. Lee, J. Tang, L. Jiang, M. Zhou, and P. Li, "User-level sentiment analysis incorporating social networks," In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, pp. 1397-1405, 2011.

[12] M. Sachan, and R. Ichise, "Using semantic information to improve link prediction results in network datasets," International Journal of Computer Theory and Engeneering, vol. 3, pp. 71-76, 2010.

[13] J. Hannon, M. Bennett, and B. Smyth, "Recommending twitter users to follow using content and collaborative filtering approaches," In: Proceedings of the fourth ACM conference on Recommender systems, Barcelona, pp. 199-206, 2010.

[14] J. Valverde-Rebaza, and A. de Andrade Lopes, "Structural link prediction using community information on twitter," In: 2012 Fourth International Conference on Computational Aspects of Social Networks,, Chongqing, pp. 132-137, 2012.

[15] Clauset A, Moore C, Newman M E J. Hierarchical structure and the prediction of missing links in networks[J]. Nature, 453(7191): 98-101, 2008.

[16] Neville J. Statistical models and analysis techniques for learning in relational data[D]. University of Massachusetts Amherst, 2006.

[17] D. Heckerman, C. Meek, and D. Koller, "Probabilistic entity-relationship models, PRMs, and plate models," In: Proceedings of the 21st International Conference on Machine Learning, Banff, pp. 55-60, 2004.

[18] K. Yu, W. Chu, S. Yu, V. Tresp, and Z. Xu, "Stochastic relational models for discriminative link prediction," In: Advances in neural information processing systems, MIT Press, Cambridge, pp. 1553-1560, 2006.

[19] B. Liu, "Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, vol. 5(1), pp. 1-167, 2012.

[20] L. W. Ku, Y. S. Lo, and H. H. Chen, "Using polarity scores of words for sentence-level opinion extraction," In: Proceedings of NTCIR-6 workshop meeting, Tokyo, pp. 316-322, 2007.

[21] H. Hu, "Research of public opinion extraction method based on text emotional computing," University of Electronic Science and Technology of China, Chengdu, 2012.

[22] Y. Li, H. Xiao, D. Li, "Research of dynamic link prediction method based on link importance," Journal of Computer Research and Development, vol. 48, pp. 40-46, 2011.

[23] G. Yuan, P. K. Murukannaiah, Z. Zhang, and M. P. Singh, "Exploiting sentiment homophily for link prediction," In: Proceedings of the 8th ACM Conference on Recommender systems, Foster City, pp. 17-24, 2014.