Predicting Trust Relationships in Social Networks Based on WKNN

Qiuyue Zhao¹, Wanli Zuo¹, Zhongsheng Tian², Xin Wang¹, Ying Wang^{3*}

¹ College of Computer Science and Technology, Jilin University, Changchun, China.

² College of Software, Jilin University, Changchun, China.

³ College of Computer Science and Technology, Jilin University, Changchun, China; Computer Science and Engineering, Arizona State University, Tempe, AZ, USA.

* Corresponding author. Tel.: 13604413931; email: wangying2010@jlu.edu.cn Manuscript submitted January 10, 2014; accepted June 2, 2014.

Abstract: Trust relationships between user pairs play a vital role in making decisions for social network users. In reality, available explicit trust relations are often extremely sparse, therefore, inferring unknown trust relations attracts increasing attention in recent years. In this paper, a new approach originating from machine learning is proposed to predict trust relationships in social networks by exploring an improved k-nearest neighbor algorithm based on distance weight (WKNN). Firstly, we extract three critical attributes from users' personal profiles and interactive information; then, an improved KNN algorithm named WKNN is proposed; finally, comparative analysis between them is performed by using real-world dataset from Epinions to evaluate their performance in trust prediction. Empirical evaluation demonstrates that the proposed framework (WKNN model) is feasible and effective in predicting trust relationships.

Key words: Trust relationships, social networks, WKNN, KNN.

1. Introduction

With the popularity of online social networks and product-recommending sites, more and more users construct their profiles online and explicitly maintain the relationships with others by social networks. It is possible and common to interact with unfamiliar people; therefore, it is often a question "Should I trust the person?" [1]. Since trust relationships constitutes the basis of interaction and transaction between users, inferring implied trust relations attracts more and more attention in recent years.

Previous work regarding trust inference mainly relies on trust propagation techniques, which is based on the established trust relationships and the structure of social networks [2]. However, the available web of trust is too sparse [3] to apply propagation techniques. In light of these, we suggest

to predict trust relationships between users more accurately by utilizing both user personal data from user's profile and user interaction data from rating information. Moreover, we observed that a user trusts another user can be described by trustee's good reputation and credibility, or their common interests, or good rating interactions between them. In this paper, we propose a new approach (WKNN model) to predict trust relationships. Our contributions are summarized as the following:

- 1) Demonstrating that authority similarity, interest similarity and rating similarity are three critical attributes to deduce trust relationships in social networks
- 2) Proposing a new algorithm named WKNN to model trust prediction problem

3) Presenting a meaningful finding that WKNN performed better than traditional KNN algorithm

The rest of this paper is organized as follows: Section II reviews related work. Section III describes the framework for assessing trust relationships. Section IV analyzes the attributes from user profile and interaction information. Section V introduces the improvement of KNN algorithm. The dataset and experimental results are described in Section VI. We finally conclude and present future work in Section VII.

2. Related Work

Predicting trust relationships is an important topic in social network analysis. Most of the existing work about online trust is based on the following three aspects:

The first approach is based on machine learning. Xiang *et al.* [4] develop an unsupervised model to estimate relationship strength from interaction activity [5,] [6] (e.g., communication, tagging) and user similarity with the goal of automatically distinguishing strong relationships from weak ones. There are several works on inferring trust relationships by considering features such as different categories: attribute similarity, topological connectivity, transactional connectivity, and network-transactional connectivity, or by considering critical aspects of social trust consisting of reputation, knowledge, similarity and personality-based trust [7]-[10]. Zhao *et al.* [6] proposed a general framework to measure relationship strengths between different users, taking consideration not only user's profile information but also interaction activities and activity fields. Ma *et al.* [11] resolve trust prediction problem based on personalized and cluster-based classification methods. However, there are some other machine learning methods for future study.

The second approach is based on probability and statistics. Achim *et al.* [12] shown how to implement and learn context-sensitive trust using statistical relational learning in form of a Dirichlet process mixture model called Infinite Hidden Relational Trust Model (IHRTM). A new trust inference algorithm that uses probabilistic sampling to separately estimate trust information is presented in SUNNY [13]. Developing a framework for classifying the type of social relationships by learning across heterogeneous networks attracts more and more attention. A framework incorporating social theories into a factor graph model effectively improves the accuracy of inferring the type of social relationships in a target network [14]. In order to establish trust relationships between Identity providers (IdPs) in different cycle of relationship between visited IdP and home IdP, Chen *et al.* [15] design an effective authentication model that can fully use existing interface without additional negotiation processes. Crandall *et al.* [16] investigate the problem of inferring friendship between people from co-occurrence in time and space by probabilistic models.

The last approach is based on network structure and semantic reasoning. Zhang *et al.* [17] propose a semantic-based trust reasoning mechanism to mine trust relationships from online social networks automatically and exploit role-based and behavior-based reasoning functions to infer implicit trust relationships and category-specific trust relationships. Giacomo *et al.* [18] decompose a trust network into its ego network components and mine this ego network setting to find trust relationships. As we know, in product review sites people place trust differently on different people from different fields. Tang *et al.* [19] propose a novel approach by incorporating these multi-faceted trust relationships into traditional rating prediction algorithms to reliably estimate their strengths. In Financial Services Information Systems (FSIS) the trust relationship plays an important role as it has a positive effect on customer satisfaction. Certain system properties of FSIS create unique challenges for building trust relationships with the customers. To tackle these challenges, a trust topology consisting of reputation, experience and communication to describe the dynamics of the trust relationship is proposed [20]. A trust evaluation model based on fuzzy theory in P2P networks is proposed [21].

3. Framework for Predicting Trust Relationships

Trust between users is an important piece of knowledge that can be exploited in search and recommendation. Therefore, it is necessary to achieve a reasonable way of predicting trust relationships. We propose a novel approach WKNN for inferring trust in social networks combined authority similarity, interest similarity and rating similarity. The whole process includes getting attributes, analyzing attributes and constructing model, shown in Fig. 1. Through WKNN model, the relationship between users could be classified into trust or distrust.

4. Attributes Analysis

Epinions is a real-world online product review community where the public can sign up as a member. Members can rate items by writing reviews and establish trust networks with their like-minded users. Everyone can create their own personal home page that includes total visits, member visits, review rating, product rating, member trust and trusted, review written, etc. According to the information from personal home page and interaction, we clarify them as three attributes, authority similarity, interest similarity, and rating similarity.



Fig. 1. Framework for predicting trust relationships.

4.1. Authority Similarity

As we known, the more authoritative the user is, the more creditable the user will be. Furthermore, if the authority similarity between two users is greater, the more possible they establish trust relationships. Hence it is reasonable that we regard authority similarity as an attribute. We find that the credibility and visibility based on improved PageRank can obtain the authority similarity, thus we focus on them below.

User credibility depends on the user's ratings for a given item based on past ratings of other information, and if the review rating is high, people would think his credibility may not be low. The credibility can be formulated as,

$$Cred(u_i) = \alpha \frac{num_{u_i}(very)}{num_{u_i}(total)} + (1-\alpha) \frac{num_{u_i}(most)}{num_{u_i}(total)}$$
(1)

where $Cred(u_i)$ is the credibility of user u_i . $num_{ui}(very)$ is the number of the review rating as "very helpful" of user u_i , num_{ui} (most) is the number of the review rating as "most helpful" of user u_i and num_{ui} (total) is the whole number of the review rating of user u_i . Parameter α controls the contributions of these two parts.

User visibility is determined by visit quantity, which includes total visits, member and non-member visits. Generally, a large number of visits indicates high visibility, thus the discussion of visits is imperative. User visibility can be formulated as,

$$Visi(u_i) = \beta \frac{mem(u_i)}{total(u_i)} + (1 - \beta) \frac{unmem(u_i)}{total(u_i)}$$
(2)

where $Visi(u_i)$ represents the visibility of user u_i , $mem(u_i)$ is the member visits of user u_i , $unmem(u_i)$ is the non-member visits of user u_i and $total(u_i)$ is the total visits of user u_i . Parameter β controls the contributions of these two parts.

According to user credibility and visibility, we compute user's authority similarity in combination of improved PageRank algorithm (Im-PageRank). We define the Attributes Probability Matrix (A), the Initial Probability Distribution Vector (P_i), the Unit Column Vector E and the authority similarity between users u_i and u_j (*Auth*(*i*, *j*)) as following:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \end{bmatrix}$$
$$P_i = (p_i(1), p_i(2))$$
$$E = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$$
$$PageRank(u_i) = (1-d) + d \frac{P_i \cdot A \cdot E}{n}$$
(3)

$$Auth(i, j) = \frac{PageRank(u_i)}{PageRank(u_i)}$$
(4)

where both a_{1i} and $p_i(1)$ are the credibility of user u_i , a_{2i} and $p_i(2)$ are the visibility of user u_i , parameter d is a damping factor, and PageRank(u_i) is the authority of user u_i .

4.2. Interest Similarity

Generally, users tend to connect to other users due to similar preferences they share, often reflected in similar interests. So we need to devise a mechanism that can assess the interest similarity among social group members. Interest is presented by items they share and review. Therefore interest similarity between users u_i and u_j can be measured by *SimInter*(i, j):

$$SimInter(i, j) = \frac{num_C}{num_{Cate}(u_i) + num_{Cate}(u_j) - num_C}$$
(5)

$$num_{c} = num(Cate(u_{i}) \cap Cate(u_{i}))$$
(6)

Volume 10, Number 1, January 2015

where $num_{Cate}(u_i)$ and $numCate(u_j)$ refer to the categories of user u_i interested in and user u_j interested in respectively; num_C is the common categories between users u_i and u_j .

4.3. Rating Similarity

In terms of rating, we consider the review on rating, the rating score given to the review, the time of rating, and the user who rates [22]. Rating score given to the targeted review reflects directly how the rater performs the evaluation. Also, the number of ratings could give more details about the experience of the rater. Users in product review sites are likely to prefer to product reviews provided by their trust networks.

Due to close relation between rating similarity and trust, it is desirable to understand trust or not in the context of the rating similarity. Rating in social networks is the review made by reviewer, which refers to the evaluation to the item or product. Thus the rating similarity SimRate(i, j) between users u_i and u_j is defined as,

$$SimRate(i, j) = \frac{\min(num_{rev}(u_i), num_{rev}(u_j))}{num_{rev}(u_i) + num_{rev}(u_i)}$$
(7)

where $num_{rev}(u_i)$ and $num_{rev}(u_j)$ are the number of reviews user u_i and u_j wrote.

5. Model Framework

K-nearest neighbor algorithm (KNN) is an effective, non-parametric and easily implemented method for classifying objects based on closest training examples in the feature space. However, classical KNN classification algorithm ignores weight difference between samples. To overcome this weakness, an improved version of KNN (WKNN) is proposed in this work.

5.1. KNN Classification Algorithm

The KNN algorithm predicts the category of test sample based on the *K* training samples which are the nearest neighbors to the test one, and judge it to that category with greatest category probability. No weight difference exists among samples, that is, all training samples are treated equally, with no apparent difference among the samples data. In KNN algorithm [23], the process of classifying sample *X* is shown below:

Step 1: For *n* given categories C_1 , C_2 , ..., C_n , *K* training samples are mapped onto m-dimension feature vectors.

Step 2: Map sample *X* onto be the same feature vector space of the form (X_1 , X_2 , ..., X_m), as all training samples.

Step 3: Calculate the distance between all training samples and *X*. Take the j_{th} sample d_j (d_{ji} , d_{j2} , ..., d_{jm}) as an example, the distance DIS(X, d_j) is defined as:

$$DIS(\mathbf{X}, \mathbf{d}_{j}) = \sqrt{\sum_{i=1}^{m} (X_{i} - d_{ji})^{2}}$$
(8)

Step 4: Choose *k* samples closest to *X* from *K* by $DIS(X, d_j)$, (*j*=1, 2, ..., *K*), and treat them as a KNN collection of X. Then, calculate the probability of *X* belonging to each category respectively based on the following formula:

$$f(X) = \arg\max Score(X, C_n)$$

=
$$\sum_{d_j \in KNN}^{n} DIS(X, d_j) y(d_j, C_n)$$
 (9)

where Score(X, C_n) is the score of the candidate category Cn with respect to X, $y(d_j, C_n)$ is a category attribute function, which satisfies

$$\mathbf{y}(d_{j},C_{n}) = \begin{cases} 1, & d_{j} \in C_{n} \\ 0, & d_{j} \notin C_{n} \end{cases}$$
(10)

Step 5: Determine the category of X as C_l with the largest f(X).

5.2. Improved KNN Classification Algorithm

Taking into account that noise has great impact on traditional KNN algorithm, to make the classification result more robust with lower misclassification errors, we propose a new K-nearest neighbor based on distance weight method (WKNN) that assigns different weights to different distances. The algorithm is given in Table 1.

Table 1. The Algorithm of WKNN					
Input: <i>k</i> : number of nearest neighbors;					
N : number of categories;					
<i>L_i</i> : mean of <i>k</i> distances;					
L_k : k shortest distances;					
<i>DW_k</i> : weight of distance;					
<i>D_i</i> : weighted sum of distances.					
Output: <i>n</i> : classification result					
1. begin					
2. map test samples and training samples onto the					
same feature vector space in the form of $(X_1, X_2,,$					
X_m);					
3. for (i = 1; i <= n; ++i)					
4. calculate the distance between each test vector					
and each training vector to which belongs					
category C_i ;					
5. select k shortest distances L_k ;					
6. for(each l_k in L_k)					
$sum += l_k$;					
end for					
7. get the mean of L_k ;					
8. compute the weight of distance DW_k ;					
9. $for(j=1; j \le k; j++)$					
get weighted sum of distances <i>D_i</i> ;					
end for ;					
10. end for;					
11. obtain the smallest D_n from D_i ;					
12. return n;					
13. end.					

Taking into account the inequality of samples, we put forward the weight of distance W_k formulated as:

$$W_{\rm k} = \frac{L_k - \overline{L_i}}{L_k^2} \tag{11}$$

The significant improvement of WKNN algorithm is weighting to the contributions of *k* near neighbors. Thus, in order to implement this approach, firstly we need to find the *k* neighbors which have the smallest distances to the node. Then we should compute the weights of each *k* nearest neighbors based on WKNN algorithm and obtain the distance weighted sum of different categories. Finally, according to the distance values of various categories, we could determine the classification results directly.

6. Experiments and Results Analysis

In this section, we conduct experiments to evaluate the proposed framework, WKNN. First of all, we introduce the dataset and characteristics of the data used in our experiments; afterwards, we present empirical evaluation results to assess the effectiveness of WKNN model in trust prediction.

6.1. Datasets

We conduct our experiments on Epinions [24], a large product review community supporting various types of interactions, which can be used for trust classification training and evaluation. In the community, users are classified as lead, top reviewer, advisor, popular author and common user; and users rate reviews as "Off Topic", "Not Helpful", "Somewhat Helpful", "Helpful", "Very Helpful" or "Most Helpful", which determine how prominently the review will be placed. Epinions employes 5-star system to rate items and the majority of ratings are scores of 4 and 5.

Moreover, a few users have an extremely large number of trustors and trustees, while most users have few trustors and trustees, suggesting a power law distribution that is typical in social networks. The distributions of trustees and trustors for each user are demonstrated in Fig. 2.



Fig. 2. Trustees and trustors distributions in Epinions.

Our experiment setting of the dataset is demonstrated in Table 2, where users without trust time or trusted time and the categories of item are less than two are filtered out.

Table 2. Statistics of the Datasets				
Number of Users	18948			
Number of Items	283300			
Number of Categories	27			
Number of Ratings	900037			
Number of Links	310238			

6.2. Experimental Results

The experiments are carried out and validated with four test sets of different sizes (500, 1000, 2000, 3000, which are randomly selected from the dataset). The confusion matrix [7] is a useful metric for analyzing how well the classifier can recognize tuples of different classes, as shown in Table 3. Different criteria can be used to compare the performance of different classifiers, and to further determine which of them works better on a given dataset. Precision (TP/(TP+FP)), recall (TP/(TP+FN)), and F-measure (2*precision*recall/(precision+recall)) are the evaluation methods exploited to examine the performance at the level of class lable using confusion matrix, and the representative result of experiment based on the

three evaluation methods are shown in Table 4.

Table 3. Confusion Matrix							
	correct classification						
	TP	FP					
experiment	(True Positive)	(False Positive)					
al results							
	FN	TN					
	(False	(True Negative)					
	Negative)						

The results demonstrate that WKNN classifier outperforms KNN classifier. In the above experiment, we achieve stratified results when all the recalls of test sets are 1. Although we have set k=5 in the above experiment, it will be interesting to investigate the influence of different k values in terms of accuracy rate since the parameter k controls the contributions of experimental results. The corresponding results are shown in Fig. 3.

Table 4. Results of the Two Methods									
The size of test	Precision		Recall		F-Measure				
set	KNN	WKNN	KNN	WKNN	KNN	WKNN			
500	0.7964	0.9091	0.7200	1	0.7563	0.9524			
1000	0.7450	0.8498	0.6998	1	0.7216	0.9187			
2000	0.7915	0.8569	0.7970	1	0.7942	0.9229			
3000	0.7597	0.8432	0.7733	1	0.7665	0.9149			



Fig. 3. Performance of WKNN and KNN on Epinions.

Furthermore, we focus on the discussion of the test results of WKNN. In the four test sets, the sizes are set with different values. Fig. 4 shows the relationships of 500 different pairs of users which are represented by the green lines. The red nodes represent the results of relationship which are contrary to the true situation. The accuracy rate for the results reaches 95%. Fig. 5 shows the results of 1000 pairs of relationships. However, being different from the former case, the 1000 pairs of relationship reproduces the lower accuracy rate of 88.8%, and there are 112 pairs of relationships with contrary to the true situation. In addition, we show the relationships of 2000 pairs of users in Fig. 6. In this case, 167 pairs of relationships are contrary to the true situation. The accuracy rate reaches 92%. Fig. 7 shows the relationships of 3000 different pairs of users, this case gives rise to the largest discrepancy (329) between our experimental results and true situation. In this case, the accuracy rate is 89%. Based upon the analysis of these results, we find that the distance weight-based WKNN model is reasonable, which is superior to nodes-based distance KNN model in improving classification accuracy.



Fig. 4. Results of 500 pairs of relationship.



Fig. 6. Results of 2000 pairs of relationships.



Fig. 5. Results of 1000 pairs of relationships.



Fig. 7. Results of 3000 pairs of relationships.

7. Conclusions and Future Work

With the pervasive of social media, trust, as a major part of human interactions, plays an important role in social networks, which helps users collecting reliable information in social web applications. In this work, we studied the prediction of trust and distrust relationships between users in social networks. By exploiting users' profiles and interactive information, we determined three critical factors which influence online social trust, that is authority, interest, and rating. We proposed a novel framework, WKNN, which is an improved k-nearest neighbor method based on distance weight. Better findings are observed in our experiments by using real-world dataset from Epinions. We demonstrated that WKNN can be applied to significantly improve the performance of trust prediction.

There are some further research directions to be investigated. Firstly, we will continue our research on trust prediction by comparing their performances. Secondly, we will investigate trust evolution in different fields based on dynamics in the context of online product reviews. Finally, sophisticated models could be

explored to estimate the strengths of trust relationships with the aim of quantifying trust analysis result.

References

- Paolo, M., & Paolo, A. (2005). Controversial users demand local trust metrics: An experimental study on Epinions.com community. *Proceedings of the National Conference on artificial intelligence* (pp. 121-126). Pittsburgh, Pennsylvania.
- [2] Jan, O., Ellery, W., & Christopher, P. (2012). Trust propagation with mixed-effects models. *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media* (pp. 535-538).
- [3] Kim, Y. A., Le, M.-T., Lauw, H. W., Lim, *et al.* (2008). Building a web of trust without explicit trust ratings. *In Workshop on Data Engineering for Blogs, Social Medial and Web 2.0* (pp. 531-536). Mexico.
- [4] Neville, J., *et al.* (2010). Modeling relationship strength in online social networks. *Proceedings of the International Conference on World Wide Web (WWW)* (pp. 981-990). Raleigh, NC, USA.
- [5] Indika, K., & Jennifer, N. (2009). Using transactional information to predict link strength in online social networks. *Proceedings of the Third International ICWSM Conference* (pp. 74-81).
- [6] Jin, Y., *et al.* (2012). Relationship strength estimation for online social networks with the study on facebook. *Neurocomputing*, *95(15)*, 89-97.
- [7] Kiyana, Z., & Abdollah, A. (2010). Mining trust and distrust relationships in social web applications. *Institute of Electrical and Electronics Engineers IEEE* (pp. 73-80). Austin, Texas, USA.
- [8] Leskovec, J., Huttenlocher, D., & Kleinberg, J. (2010). Predicting positive and negative links in online social networks. *In WWW'10* (pp. 641–650). Raleigh, North Carolina, USA.
- [9] Lin, G. Y., Bie, Y. Y., & Lei, M. (2013). Trust based access control policy in multi-domain of cloud computing. *Journal of Computers*, *8*(*5*), 1356-1365.
- [10] Han, T., Luo, S. S., Zhu, H. L., Xin, Y., & Peng, Y. (2013). A novel trust evaluation model based on grey clustering theory for routing networks. *Journal of Computers*, *8*(*11*), 2743-2749.
- [11] Nan, M., et al. (2009). Trust relationship prediction using online product review data. Proceedings of the 1st ACM International Workshop on Complex Networks Meet Information & Knowledge Management (pp. 47-54). New York, USA.
- [12] Achim, R., Matthias, N., & Volker, T. (2011). Statistical relational learning of trust. *Machine Learning*, *82(2)*, 191-209.
- [13] Ugur, K., & Jennifer, G. (2010). Using probabilistic confidence models for trust inference in web-based social networks. *ACM Transations on Internet Technology*, *10(2)*, pp. 1-23.
- [14] Jie, T., *et al.* (2012). Inferring social ties across heterogeneous networks. *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining* (pp. 743-752). New York, USA.
- [15] Wang, Y.-H., et al. (2010). Trust relationship establishment based on the existing trust. 2010 Second International Workshop on Education Technology and Computer Science (ETCS) (pp. 141-144). Wuhan, China.
- [16] Crandall, D. J., Backstrom, L., *et al.* (2010). Inferring social ties from geographic coincidences. *PNAS*, *107*, 22436–22441,
- [17] Yu, Z., & Tong, Y. (2012). Mining Trust relationships from online social networks. *Journal of Computer Science and Technology*, *27*(*3*), 492-505
- [18] Giacomo, B., Michele, C., Anna, M., & Fosca, G. (2012). Classifying trust/distrust relationships in online social networks. *Proceedings of IEEE International Conference on Social Computing* (pp. 552-557). Amsterdam, Netherlands.
- [19] Huan, L. *et al.* (2012). mTrust: Discerning multi-faceted trust in a connected world. *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining* (pp. 93-102). New York, USA.

- [20] Fengwei, Y. (2010). Managing the trust relationship in financial services information systems. *Proceedings of 2010 2nd IEEE International Conference on Information and Financial Engineering(ICIFE)* (pp. 207-211). Chongqing, China.
- [21] Zhou, Z. Z., *et al.* (2011). A trust evaluation model based on fuzzy theory in P2P networks. *Journal of Computers (JCP)*, *6(8)*, 1634-1638
- [22] Jaideep, S., *et al.* (2008). Predicting trusts among users of online communities-an Epinions case study. *Proceedings of the 9th ACM Conference on Electronic Commerce* (pp. 310-319), New York, NY, USA.
- [23] Suguna, N., *et al.* (2010). An improved k-nearest neighbor classification using genetic algorithm. *International Journal of Computer Science Issues*, 7(4).
- [24] Huan, L., et al. (2012). eTrust: Understanding trust evolution in an online world. Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 253-261). New York, USA.



Qiuyue Zhao was born in China in 1989. She is a master student majoring in computer science at Jilin University, Changchun, China. Her main research interests include social networks, date mining and machine learning.



Wanli Zuo was born in China in 1957. He received a Ph.D in College of Computer Science and Technology of Jilin University. He is currently a professor at Jilin University, Changchun, China. His main research interests include database, date mining, machine learning and search engine.



Zhongsheng Tian was born in China in 1989. He is a master student majoring in computer software at Jilin University. His main research interests include social networks, date mining and machine learning.



Xin Wang was born in China in 1981. He is a Ph.D candidate in College of Computer Science and Technology of Jilin University. His main research interests include web information retrieval and mining, deep web and machine learning.



Ying Wang was born in China in 1981. She received her Ph.D degree from College of Computer Science and Technology, Jilin University. She is currently a lecturer at Jilin University and a visiting scholar at Computer Science and Engineering Department of Arizona State University. Her main research interests include social networks, date mining, machine learning and search engine.