International Journal of Hybrid Information Technology
Vol.8, No.12 (2015), pp. 205-214
http://dx.doi.org/10.14257/ijhit.2015.8.12.14

The Research of Recommendation Algorithm based on Complete Tripartite Graph Model

Wu Pin, Liao Shiwei, Lv Bo, Zhu Yonghua, Gao Honghao and Lin Ju

School of Computer Engineering and Science, Shanghai University, 200444, Shanghai, China
Institute of Materials Genome, Shanghai University, Shanghai, China
Computing Center, Shanghai University, 200444, Shanghai, China
wupin@shu.edu.com.cn

Abstract

Social tags providing abundant information can stimulate a better recommender system equipped with stronger sense of description and analysis on user's interest. In this paper, graph-based personalized recommendation techniques have been studied. Complete tripartite graph model was proposed and the user's interest migration was researched comprehensively, Focusing on the dilemma of accuracy and diversity in recommender system, the mass diffusion algorithm and heat spreading algorithm on complete tripartite graph model were carried out. Then, from the perspective of improving confidence in recommender system, the item-tag joint recommendation mechanism was studied. Experimental results show the effectiveness of the algorithm in this paper.

Keywords*: social tags, personalized recommendation, complete tripartite graph*

1. Introduction

In recent years, with the continuous development of the Internet and Web 2.0, we encounter the information era. Users now undergo even more pains and sweat to filter for what they need out of massive data. Personalized recommendation [1] reveals user's interest in their past behaviors, and thus recommend for them. Social tagging [2] is a symbolic technique today powering up websites with the freedom that users can label information by their preferences, which will in the long run, turn into recommendation in return.

At present, tag models proposed by scholars can be divided into three kinds: model based on probability, model based on tripartite graph and model based on tensor. Many scholars have also put forward recommendation algorithm considering social tags. For example, paper [3] proposed user-centered similarity based on physical diffusion to obtain more accurate recommend results; paper [4] directly saw tag's use frequency as the edge's weight value, and used the diffusion algorithm to improve the accuracy; paper [5] considered tag-used model, and adopted the TF - IDF model to calculate the weight value of User-Item relations. This paper fully considers the relationship between users, items and tags, proposing a user-item-tag complete tripartite graph model to improve the accuracy of recommendation.

2. Complete Tripartite Graph Model

Some scholars put forward introducing social tagging into collaborative filtering based on the graph, in order to improve the recommendation results accuracy and interpretability. Tag as a kind of independent node, added into user-item binary diagram, user-item-tag tripartite graph model is formed, shown in Figure 1 [6]. In this paper, using graph G (V, E) to describe tripartite graph structure, which contains three kinds of node:

ISSN: 1738-9968 IJHIT

user node (VU)、item node(VI) and tag node(VT)，and $V = V_U \cup V_I \cup V_T$.

Most recommendation algorithms previously understand the relationships between three kinds of node into two bipartite graphs' relationship, losting the other connection information in the graph model, and therefore we can use edge relationship to connect two different kinds of nodes, to reflect three kinds of entities' relationship. This paper presents complete tripartite graph model considering the relationship between the three kinds of node, As is shown in the Figure 2, we can see, the relationship between the different nodes are reflected through the edges.

We are able to build the user - item relation matrix (U-I matrix), item-tag relation matrix (I-T matrix) and user - tag relation matrix (U-T matrix) to discribe the connection relationship:

U - I matrix(B_{UI}): if a user U_i has selected a item I_j, then b_{ij}=1,else b_{ij}=0;

I - T matrix(B_{IT}): if an item I_j is labeled by a tag T_l， then b'_{jl}=1, else b'_{jl}=0;

U - T matrix(B_{UT}): if a user U_i has used a tag T_l, then b''_{il}=1,else b''_{il}=0;

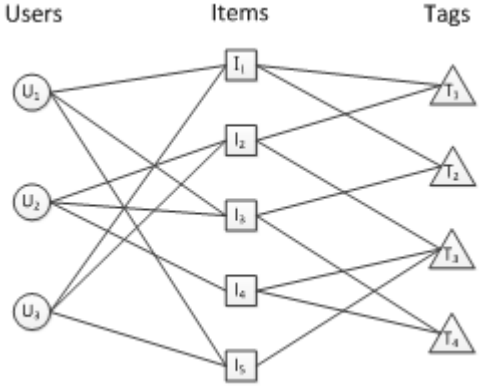

Figure 1. Tripartite Graph Model

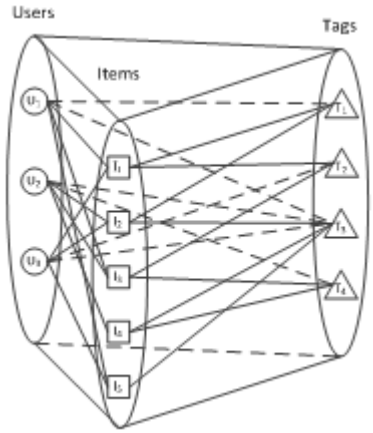

Figure 2. Complete Tripartite Graph Model

3. Recommendation Algorithm

3.1. The Analysis of User's Interest Migration

In the real network, information is constantly updated, the user's preferences are changing over time, so when recommending items to users, we should take time effect into consideration.

Generally speaking, users are most interested in items selected recently than items selected long ago. User's interest migration may exist for each user, on the one hand, the

user's interests can be divided into long-term interests and short-term interests, long-term interests has certain stability, but short-term interests ofen fluctuate, then causes the network behavior change. On the other hand, the change of personality will lead to the change of their network behavior.

In the above models, the connection relationship between the nodes in the diagram are equal, for example, elements only represente 1 or 0. The relationships between nodes have only two kinds: have connection and have no connection. There will be two problems with this setting: firstly, Can't reflect the user's interest in migration; secondly, it can't reflect the degree of users interested in different items.

So, this paper distributes weight for nodes in the complete tripartite graph model according to the time and the number of user's behavior, take user U_i and item I_j and define the connection weight between two nodes in formula(1).

$$w_{ij} = \sum_{s=1}^{k} \left(\frac{w_0}{1 + e^{(t-t_s)t_0}} + w_0 \right) \tag{1}$$

In this formula, t is current time and k is the number of same behavior. ts is the time when user U_i's behavior to item I_j occurs. t_0 is user's interest migration's time factor, w0 representes weight threshold, namely, with time going on, recommended capacity provided by user's behavior will be able to gradually decreases, and finally tends to a constant, it is set to 0.5.

Through the weight value and the initial matrix, we can get a new connection relationship matrix B_{UI}、B_{IT}、B_{UT}, use formulate (2) to build a new user - item relationship matrix.

$$b_{ij} = w_{ij} \times a_{ij} \tag{2}$$

In this formula, a_{ij} is corresponding element in the user - item relationship matrix.

Similarly, we can acquire the item-tag relation matrix and user - tag relation matrix, the new relation matrixs describe connection relation strength between three kinds of nodes in the complete tripartite graph model, and connection weight considers uer's interest migration, so the model with time weight can reflect the relationship between user, tag and item better.

3.1. Complete Tripartite Graph Mass Diffusion Algorithm (CTGMD)

CTGMD is based on the idea of probability propagation i, the flow chat of CTGMD is shown by Figure 3.

First, the data set preprocessing, and build the user - item relation matrix (U-I matrix), item-tag relation matrix (I-T matrix) and user - tag relation matrix (U-T matrix).

Second, allocate initial resources. For a particular user U_i, we can allocate resources for its all adjacency item nodes, then diffuse in item-user-tag-item direction and item-tag-user-item direction:

1. Diffusdion in item-user-tag-item direction, finally gets resource redistribution vector :

$$\vec{f}_j^{'} = \sum_{i=1}^{r} \frac{b_{lj}^{''}}{\sum_{o=1}^{m} b_{lo}^{n}} \sum_{i=1}^{n} \frac{b_{il}^{'}}{\sum_{o=1}^{r} b_{lo}^{'}} \sum_{s=1}^{m} \frac{b_{is} \times f_s}{\sum_{o=1}^{m} b_{os}} \tag{3}$$

In this formula, f_s is the item's corresponding initial resources in initial resources vector. n, m, r respectively corresponds the number of user node, item node and tag node. $\sum_{o=1}^{n} b_{os}$ represents item Is's adjacent user's number.

2. Diffusion in item-tag-user-item direction, finally gets resource redistribution vector :

$$\vec{f}_j^{''} = \sum_{i=1}^{n} \frac{b_{ij}}{\sum_{o=1}^{m} b_{io}} \sum_{l=1}^{r} \frac{b_{il}^{'}}{\sum_{o=1}^{m} b_{lo}^{'}} \sum_{s=1}^{m} \frac{b_{ls}^{n} \times f_s}{\sum_{o=1}^{m} b_{os}^{''}} \tag{4}$$

Finally combining two direction diffusion redistribution's resources as a result, get the final resource vector:

$$\vec{f}^{*} = \lambda \vec{f}^{'} + (1-\lambda)\vec{f}^{''} \quad (5)$$

In this formula, $\lambda \in [0,1]$, when λ=0, initial resources diffuse in item-tag-user-item direction. While, λ=1, initial resources diffuse in item-tag-user-item direction and we can finally gets resource redistribution vector. We can also adjust the value of λ, to adjust the proportion of two direction mess diffusion.

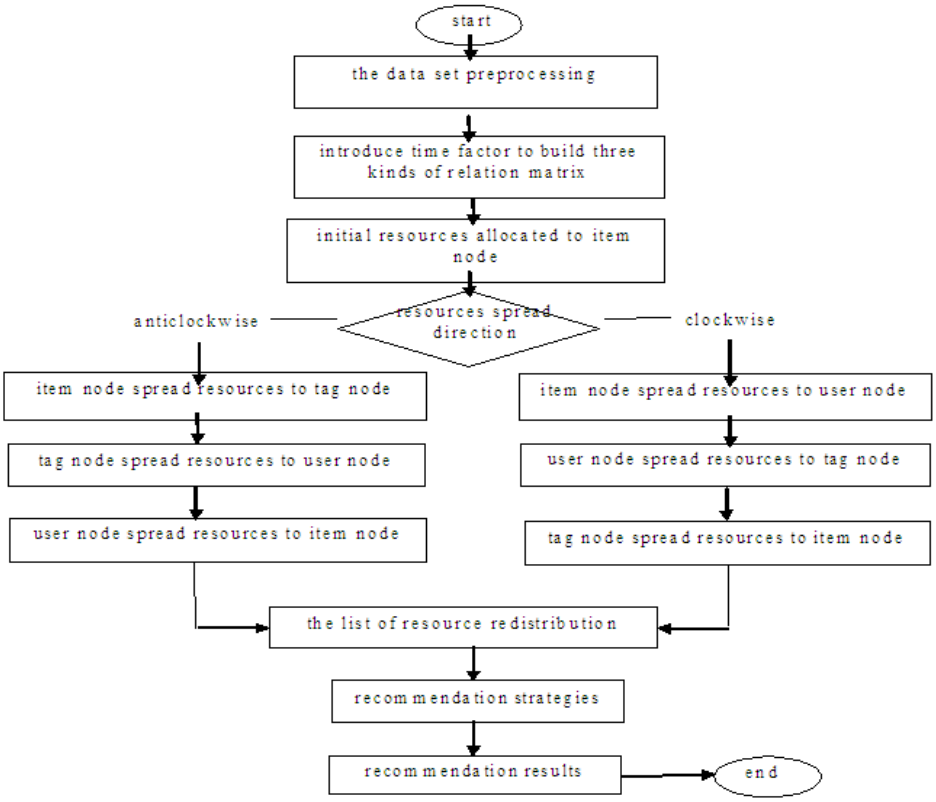

Figure 3. CTGMD Flowchart

3.2. Complete Tripartite Graphs Heat Spreading Algorithm (CTGHS)

From the paper [7], we can know, using Heat Spreading algorithm to allocate resources makes recommendation result's diversity better. MassDiffusion algorithm's thought is that resources from one node according to the number of adjacent nodes, evenly spread to another node. However, Heat Spreading algorithm's thought is that a node from adjacent nodes according to the number of adjacent nodes, evenly absorb resources.

In the early stages of the algorithm, CTGHS algorithm is similar to CTMGD algorithm. Namely, data needs once preprocessing, and Obtain by calculating U-I matrix, I-T matrix and U-T matrix. For a particular user U_i, we can allocate resources for its all adjacency item nodes, and then absorb heat in user-tag-items direction and items - tag - user direction.

1. Absorbing heat in Item-User-Tag-Item direction, finally gets resource redistribution vector :

$$\vec{f}_{j}^{'} = \sum_{i=1}^{r} \frac{b_{lj}^{''}}{\sum_{0=1}^{r} b_{jo}^{''}} \sum_{i=1}^{n} \frac{b_{il}^{'}}{\sum_{0=1}^{n} b_{lo}^{'}} \sum_{s=1}^{m} \frac{b_{is} \times f_s}{\sum_{0=1}^{m} b_{io}} \quad (6)$$

2. Absorbing heat in Item-Tag-User-Item direction, finally gets resource redistribution vector :

$$\vec{f}_j^{''} = \sum_{i=1}^{n} \frac{b_{ij}}{\sum_{0=1}^{n} b_{jo}} \sum_{i=1}^{r} \frac{b_{il}^{'}}{\sum_{0=1}^{r} b_{lo}^{'}} \sum_{s=1}^{m} \frac{b_{ls}^{''} \times f_s}{\sum_{0=1}^{m} b_{lo}^{''}} \quad (7)$$

The last resource redistribution results can be calculated using formula (5).

3.3. Hybrid Recommendation Algorithm

Mess diffusion [8] can obtain better accuracy, and heat transmission can obtain better diversity, while they contradict each other, The general solution is to synthesize better accuracy's and diversity's algorithm. As formula (8) shows, this paper synthesize, CTGHS algorithm and CTMGD algorithm by the way of Linear Mixture.

$$\vec{f}^{''} = u \frac{f_M^*}{|\vec{f}_M^*|} + (1-u) \frac{f_H^*}{|\vec{f}_H^*|} \quad (8)$$

In this formula, $\vec{f}_M^*$ is resource vector though CTGMD algorithm, $\vec{f}_H^*$ is resource vector though CTGHS algorithm. u $\in$ [0, 1] is mixed proportion adjustment factor, when u=0, only use CTGHS algorithm and when u=1, only use CTMGD algorithm. Becauese resource is not conservative in CTGHS algorithm, so hybrid recommendation algorithm need consider normalization.

4. Items-tag Joint Recommendation Mechanism

Item-tag joint recommendation mechanism consists of two stages: first, use the user's data of historic behavior to build recommendation model; second, recommend tags.

Recommendation system [9] build recommendation mode by user's data of historic behavior. Running relevant algorithm on recommendation mode gets recommendation results. Then in view of the recommended items, run relevant algorithm on recommendation mode gets the corresponding tag recommendation results. Whether item recommendation or tag recommendation results, when presented to the user, the user will produce certain feedback to its, such as Choosing recommended items, or using the corresponding tags label recommend items. User's feedback can be used to update recommendation mode, thus improve recommendation results of items and tags. Figure 4 shows the recommendation mechanism

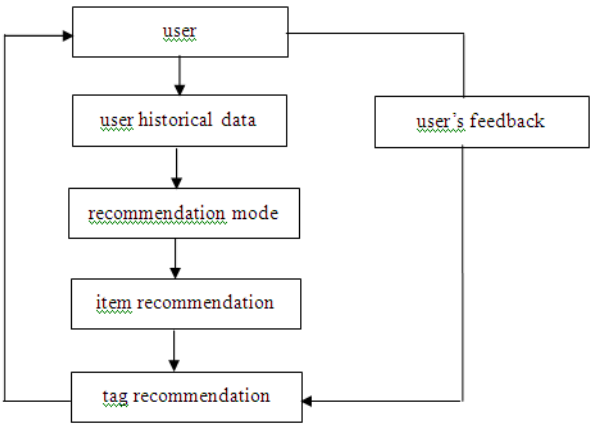

Figure 4. Items-tag Joint Recommendation Mechanism

5. Experimental Results and Analysis

MovieLens is a virtual community web site, which is also a recommendation system. It allows users to use tags to label the film they have seen. Tags can be a movie theme, actor's last name, the year of release, *etc*. This paper uses 10M datasets from MovieLens research team. Because lengths of tags are different, some of them even have no specific semantics, so we need to filtrate tags to simplify data. In order to ensure that each node has connection with the other two kinds of nodes, we must remove the independent node objects. Preprocessing results are shown in Table 1.

Table 1. MovieLens 10 Mdatasets Preprocessing Results

	MovieLens 10M datasets
Total number of users	2766
Total number of films	4758
Total number of tags	4153
Total number of score's data	18873
The least estimated number	20
The least used number of tags	50
User minimum life cycle	a month

After preprocessed, the datasets are divided into training set and testing set according to certain proportion, The former is thought as the known information of algorithm implementation, the latter is used to test the algorithm performance.

A suitable evaluation index is the key to measure recommendation algorithm, the indexes adopts in this paper adopts are precision rate, recall rate, hitting rate and diversity.

1. Precision rate: namely the ratio of the number of the items satisfied and the items in recommendation list, precision rate reflects current user's level of satisfaction on results recommended .Formal definition is as following:

$$precison = \frac{\sum_{u \in U} | R(u) \cap T(u) |}{\sum_{u \in U} | R(u) |} \quad (9)$$

R(u) is recommendation list given by recommendation algorithm according to user's conduct on training set;T(u) is user's conduct list on testing set.

2. Recall rate: the ratio of the number of items satisfied and items loved by user in recommender system, recall rate reflects the probability that user find fond items by recommendation algorithm. Formal definition is as following:

$$\text{Re } call = \frac{\sum_{u \in U} | R(u) \cap T(u) |}{\sum_{u \in U} | T(u) |} \quad (10)$$

3. Hitting Rate [10]: the ratio of the number of items satisfied and the length of recommendation list. When the item chosen by user is in user's recommendation list, algorithm hits this record. Hitting rate depends on Top N recommendation's length. Formal definition is as following:

Hitting=number of items satisfied / the length of recommendation list (11)

4. Diversity [11]: Zhoutao *et al.*, Proposed that hamming distance can be adapted to measure list's diversity. Hamming distance of User U_i's and U_j's recommendation results is as follows:

$$H_{ij} = 1 - Q_{ij} / L \quad (12)$$

Q_{ij} is the number of that U_i's and U_j's recommendation results are same; L is the length of recommendation list. Generally speaking, the bigger hamming distance is, the better diversity is.

In order to test Complete Tripartite Graphs Hybrid recommendation algorithm (CTGH),we can compare it with Item-Based Collaborative Filtering algorithm(IBCF) [12] and Integrated Diffusion on Tripartite Graphs algorithm (IDTG) [13], three algorithm adopt Top-N recommendation to conduct this experimenting. In CTGH algorithm, t_0=14, λ=0.6，μ=0.2. Tabel 2, Table 3, Table 4 Lists the contrast data respectively:

Table 2. Precision Rate

Recommendation list length	IBCF	IDTG	CTGH
10	0.0632	0.0861	01125
50	0.0561	0.0645	0.0697
100	0.0343	0.0487	0.0563

Table 3. Recall Rate

Recommendation list length	IBCF	IDTG	CTGH
10	0.0132	0.0163	0.0254
50	0.0275	0.0318	0.0422
100	0.0352	0.0435	0.0492

Table 4. Recommendation List Length

Recommendation list length	IBCF(%)	IDTG(%)	CTGH(%)
10	14.5	17.1	19.1
50	37.9	43.2	45.5
100	51.9	57.1	59.8

Table 5. Diversity

Recommendation list length	IBCF	IDTG	CTGH
10	0. 65443	0. 89323	0.90829
50	0. 50274	0. 81374	0.82015
100	0. 43848	0. 73385	0.75172

Experiments show that CTGH algorithm is better than IBCF and IDTG algorithm in precision rate, recall rate, hitting rate and diversity. So Complete Tripartite Graphs Model put forward in this paper can better reflect the users' interest, on the basis of which, item recommendation can get better effectiveness.

6. Conclusion

This paper mainly studies the personalized recommendation technology, put forward Complete Tripartite Graphs Model and corresponding algorithm, experimental results show the effectiveness of the algorithm in this paper. With the development of the recommendation system, I will work in more research direction: (1) Improving the quality of tag. In this paper, pretreatment of the datasets don't completely delete meaningless or useless tags, at the same time, tags may have a number of different meaning, which will

make recommended results inaccurate. (2) Cold start and data sparseness problem. These papers don't consider the two problem, which will be important direction of future research. (3) Big data and incremental calculation problem. User's information grow continuously and fastly, that how to make use of big data to improve the recommendation performance has significant value.

Acknowledgments

This paper is supported by the research grant (No.14DZ2261200) from Shanghai Government and Shanghai Institute of Materials Genome, Foundation of Science and Technology Commission of Shanghai Municipality under Grant No. 14590500500, Natural Science Foundation of Shanghai under Grant No. 15ZR1415200, and Young University Teachers Training Plan of Shanghai Municipality under Grant No. ZZSD13008. I would like to express my gratitude to all those who helped me during the writing of this paper. I gratefully acknowledge the help of my supervisor ZHU Yonghua, I do appreciate her patience, encouragement, and professional instructions during my paper writing. Also I would like to thank Mr Zhu Yonghua, who kindly gave me a hand when I encountered difficulties. In addition, I deeply appreciate the contribution to this paper made in various ways by my friends.

References

[1] G. Adomavicius and A. Tuzhilin, "Towards the next generation of recommender systems：A survey of the state-of-the-art and possible extensions", IEEE Trans on Knowledge and Data Engingeering, vol. 17, no. 6, **(2005)**, pp. 734-749.
[2] P. Lamere, "Social tagging and music information retrieval", Journal of New Music Research, vol. 37, no. 2, **(2008)**, pp. 101-114.
[3] M. S. Shang, Z. K. Zhang, T. Zhou, *et al.*, "Collaborative filtering with diffusion-based similarity on tripartite graphs", Physica A: Statistical Mechanics and its Applications, vol. 389, no. 6, **(2010)**, pp. 1259-1264.
[4] M.-S. Shang and Z.-K. Zhang, "Diffusion-Based Recommendation in collaborative Tagging Systems", Chin.Phys.Lett., vol. 26, **(2009)**, pp.11.
[5] P. Wuand and Z.-K. Zhang, "Enhancing personalized recommendation in weighted social tagging networks", Physical Procdia, vol. 3, **(2010)**, pp. 1877–1885.
[6] Z. K. Zhang, T. Zhou, Y. C. Zhang, "Personalized recommendation via integrated diffusion on user–item–tag tripartite graphs", Physica A: Statistical Mechanics and its Applications, vol. 389, no. 1, **(2010)**, pp. 179-186.
[7] T. Zhou, Z. Kuscsik, J. G. Liu, *et al.*, "Solving the apparent diversity-accuracy dilemma of recommender systems", Proceedings of the National Academy of Sciences, vol. 107, no. 10, **(2010)**, pp. 4511-4515.
[8] Y. C. Zhang, M. Medo, J. Ren, *et al.*, "Recommendation model based on opinion diffusion", Europhys Lett, vol. 80, **(2007)**, pp. 68003.
[9] P. Resnick and H. R. Varian, "Recommender systems", Communications of the ACM, vol. 40, no. 3, **(1997)**, pp. 56-58.
[10] Z. K. Zhang, T. Zhou and Y. C. Zhang, "Personalized recommendation via integrated diffusion on user –item–tag tripartite graphs", Physica A: Statistical Mechanics and its Applications, vol. 389, no. 1, **(2010)**, pp. 179-186.
[11] T. Zhou, L. L. Jiang, R. Q. Su, *et al.*, "Effect of initial configuration on network-based recommendation", EPL (Europhysics Letters), vol. 81, no. 5, **(2008)**, pp. 58004.
[12] B. Sarwar, G. Karypis, J. Konstan, *et al.*, "Item-based collaborative filtering recommendation algorithms", Proceedings of the 10th international conference on World Wide Web. ACM, **(2001)**, pp. 285-295.
[13] T. Zhou, J. Ren, M. Medo, *et al.*, "Bipartite network projection and personal recommendation", Physical Review E, vol. 76, no. 4, **(2007)**, pp. 046115.

Authors

Pin Wu, vice Professor. She received the B.S. and Ph.D. at Nanjing University of Science and Technology in 1998 and 2003. She had worked in Zhejiang University as a post doctor for two years, and had worked in Michigan State University as a senior visiting scholar for one year. She is working in the School of Computer Engineering and Science of Shanghai University now. Her current research interests are focusing on image processing, high performance computing (HPC), computational fluid dynamics (CFD) and so on. Over the past ten years, she has published over 30 technical papers in the related fields. She will keep on the research work addressing Cross-disciplinary of computer and mechanics.

Yonghua Zhu, vice Professor. He received the B.S. at Xi'an Jiaotong University in 1990, the M.S. at Tongji University in 1993 and Ph.D. at Shanghai University. He is working in the School of Computer Engineering and Science of Shanghai University. His current research interests are focusing on high performance computing, network computing, and interconnected network design, Communication and Information Engineering, Intelligent Controlling. Over the past ten years, he has published over 20 technical papers in the related fields. He will keep on the research work addressing Cross-disciplinary of communication, computer and automation.

Honghao Gao, received the Ph.D degree in the School of Computer Engineering and Science of Shanghai University, Shanghai, China, in 2012. His research interests include Web service and model checking.

