

Multi-granularity Recommendation Based on Ontology User Model

Jianxing Zheng, Bofeng Zhang, Guobing Zou

School of Computer Engineering & Science
Shanghai University
Shanghai, China
e-mail: bfzhang@shu.edu.cn

Abstract—The traditional personalized recommendation system supplies the target user with top k items in fixed interest subject. However, the recommended items cover the coarse subject level and the accuracy performance is poor. Taking into account ontology structure of subject, user's actual interests can distribute in multiple sub-subject structures. In this paper, multi-granularity recommendation mechanism relying on multi-granularity similarity is proposed to fit user's actual detail demands. Specially, a personalized ontology user model is learned to represent user's multi-granularity interests. According to ontology structure, the multi-granularity similarity method is implemented by combing content closeness and semantic closeness between user models at different grained subjects. Lastly, recommendation method distributed in multi-granularity subjects is achieved to compare against traditional single subject's recommendation for their performances. The experimental results show that the proposed mechanism is more successful.

Keywords- user model; ontology; multi-granularity; recommendation

I. INTRODUCTION

On the last decades, social media, such as Facebook, Twitter, and Sina Weibo, have flourished and raised much attention, which generate large-scale massive data. Various kinds of online recommender systems have proven to be a critical way for satisfying users' needs and improving users' reputation. However, the effectiveness of recommender systems highly relies on user model constructing and social relationship discovering. In recommender systems, personal tastes makes recommender systems identify further interests for user; semantic interest can partition various kinds of interest from perspective of fine grain level to identify friend group. However, most researches focus on user model constructing by taking one of factors into consideration. User models describing subject interest with a single interest degree measure interest of user at the aspect of coarse category in terms of social interaction. The relationships between user models are ordinary, which could not reflect the complex social links between users. Therefore, the existent similarity calculation methods between users have a problem of discovering close fine-granularity users.

Nowadays, one of the most popular social media websites, Sina Weibo, has supplied various services such as micro-blogging recommendation, topic recommendation, advertisement marketing and so on. Also, they need to obtain

a recommended list of users or products and compare them to the user model for determining what to recommend. However, there is lack of recommendation mechanism from perspective of different interest subject level which can help users to catch various grain friends and receive needs from multi-granularity levels of demand. As for an individual user, recommender systems should provide one's similar friends and appropriate products on different granularity levels.

To address these issues, in this paper, we design a novel ontology user model approach to solve the problem of user's demands in different granularities to recommend top k users. The proposed multi-granularity similarity method allows us to identify user's interest in different level of subjects and promotes to provide a suitable recommended list of friends who has different resonances with user in multi-granularity subjects. We validate our approach by conducting experiments in Sina Weibo Platform. Our experimental results show that the proposed approach could enhance the efficiency and effectiveness of product recommendation in terms of precision and recall metric.

The rest of the paper is organized as follows. Section II briefly reviews state of the art; in Section III, the multi-granularity recommendation framework is given; Section IV describes the multi-granularity similarity method between ontology user models; Section V discusses the experiments with detailed data and data analysis; conclusion will close the paper in Section VI.

II. STATE OF THE ART

A. User Model Representation

User model is a method with formal and conceptual language describing and specifying the semantic meanings of queries and capturing user information needs [1]. User model was defined by Zhong [2] as a series of interesting topics for a user to predict the actual thought decision and intention of behavior action. Kim [3] utilized conceptual clusters for mining and representing user interest by analyzing tagging practices of individual users with formal concept analysis. Semantic interest affects the accuracy of user model by enhancing robustness of model. A personalized ontology model based on world knowledge base and user local instance repository was proposed for knowledge representation and reasoning over user profiles [4]. Sieg et al. [5] learned ontology user profile from the Open Directory Project to specify users' preferences and interests in web

search. Leung [6] developed several concept-based user profiling methods by taking into account positive and negative preferences from search engine logs. In this paper, we study the design of semantic interest by analyzing ontology concept for realizing multi-granularity ontology user model, which is an emerging research avenue for exploiting multi-granularity recommendation services in the context of online micro-blogging platform.

B. Semantic Similarity Calculation

The issue of semantic similarity calculation has aroused much academic research and has been spotlighted for decades. Semantic similarity calculation can usually be measured by distance metric, which was categorized into two types: 1) text-based metrics [7][8][9][10], which verifies similarity between user models based on the recurrence of user's interest text; and 2) Structure-based metrics, which measures closeness of user interests determined by relationships between ontology concepts. To understand the semantic relationships between one word and relevance words, ontology category was used to make distinction between user models and identified the semantic distance of interests [11][12]. For a category tree, the distance of the shortest path between nodes was used to analyze semantic relations [4]. Liu proposed a weighted ontology-based semantic similarity method with information theory [13], and concept similarity and description similarity by the hierarchy of ontology concepts were used to examine relatedness of two words. However, all these studies considered the semantic links by vertical shortest distance of category tree and ignored semantic coverage strength of concepts from perspective of the same granular subjects.

C. Recommendation Approaches

Many of the conventional recommendation methods utilize two main techniques: content-based and collaborative filtering approaches to discover users' personal interests [14]. The content-based approach is used to create user model by analyzing user's previous interest intentions [15], while the collaborative-based approach adopts general tastes of similar users' models to predict future user's interests [16]. However, both filtering approaches obey one important assumption: users who have similar interests in the past are likely to have similar interests in the future [17]. Most of the proposed recommendation approaches suppose that populations are fully mixed: everyone has the same product need probability by following one's user model. Actually, every person makes a decision to choose appropriate ones according to different levels of products' satisfaction. Nevertheless, the unified type of recommendation on some topic may not completely suit the user's different granularity information need.

In our research, as we aim to support people diverse content products or messages with a wider range, this objective cannot be achieved by merely digging into user model constructing. For recommendation purpose, we need to find diverse kinds of similarities on different granularity subjects in order to select appropriate group models.

III. THE SYSTEM ARCHITECTURE

We proposed a novel multi-granularity recommendation service to provide similar users or messages for terminal users based on multi-granularity similarity between ontology user models considering content similarity and semantic similarity of user interest, as shown in Figure 1.

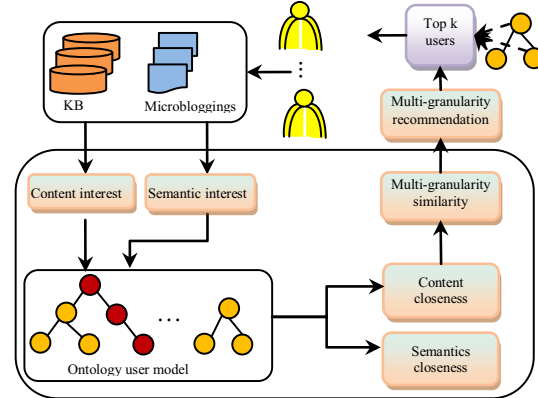


Figure 1. Multi-granularity recommendation based on multi-granularity similarity among ontology user models

In Figure 1, the first step of recommendation system is to create personalized ontology user model. The features from micro-blogging user interested in are used to match the ontology domain subjects to find user's interest subjects. Furthermore, ontology user model interest tree is built to reflect affinity levels among subject categories.

In the similarity analysis process, we also adopt a category tree structure feature to implement the content closeness and semantic closeness between different user models. For a category subject, we calculate difference of interest degree to find similar models with different similarity degree of different sub-subjects.

Finally, recommendation stage based on multi-granularity subjects' similarity selecting top k users is performed to supply terminal service. The proposed recommendation process considers the different needs of users in order to provide multi-level service from multiple aspects.

IV. THE PROPOSED MULTI-GRANULARITY SIMILARITY METHOD

A. Personalized Ontology User Model Construction

This subsection analyzes the interest degree of category subjects for a personalized ontology user model. The interest degree includes two main components: content interest degree and semantic interest degree. The occurrence times of a subject are an important factor for content interest degree. The structure location of subject in the ontology relations is a crucial factor for empowering semantic interest degree, which then is used to calculate similarities and closeness between user models.

1) Content Interest Degree of Subject

Messages posted usually contain different interest subjects. In our research, text preprocessing of stopword

removal and word stemming is processed to deal with the message, which is finally represented as $m = \{(t_1, w_1), (t_2, w_2), \dots, (t_p, w_p)\}$ formed by a set of term weighting pairs as follows:

$$w_i^m = tf_{im} * idf_i \quad (1)$$

$$tf_{im} = \frac{f_{im}}{\max_l(f_{lm})} \quad (2)$$

$$idf_i = \log \frac{N_m}{n_i} \quad (3)$$

where f_{im} represents raw frequency of term i in message m and $\max_l(f_{lm})$ stands for the frequency number of term l which has the maximum frequency in m . N_m is the number of messages in training set and n_i is the number of messages that have the same frequency number of term i as message m . Here, the term weight of a message is derived based on the $tf-idf$ method [18].

In the messages set, each message is compared with articles in four categories, and then classified into one category by *KNN* method, which was referred in previous work [19]. For a category subject s , the messages set user u involved in is represented as D_s^u . The weight of a specific subject s for user u can be finally calculated using the accumulated function:

$$w_s(u) = \sum_{m \in D_s^u} w_s^m \times \eta(s, m) \quad (4)$$

where $\eta(s, m) = 1$ if $s \in m$; otherwise $\eta(s, m) = 0$

Furthermore, content interest degree of subject s for user u is shown in Eq. (5):

$$Cid_s(u) = \frac{w_s(u)}{\max_{s' \in S_u} \{w_{s'}(u)\}} \quad (5)$$

where S_u is the set of subjects user u are interested in.

2) Semantic Interest Degree of Subject

To calculate semantic interest degree, we construct target user's personal interest tree directly by user's interest subjects. In our research, taking into account the horizontal coverage of subjects and hierarchical layer[4], semantic specificity degree of subject is designed so as to consolidate the semantic interest degree of interest category tree.

a) Semantic Specificity Degree of Subject

In this step, we will investigate the semantic specificity degree of subject according to the sibling nodes and children

nodes of subjects. The hierarchical affiliation of category "sport" is illustrated in Figure 2.

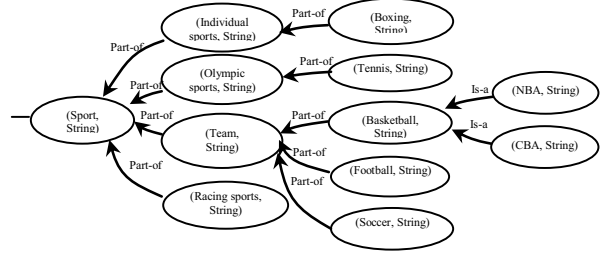


Figure 2. A sample of structure for "sport" in KB.

The detailed determination of a subject's semantic specificity degree is described in Algorithm 1. Siblings of subject s represent the semantic specificity ability of s for upper node. The subject with more siblings toward the root is more certain than that with little brothers for expressing root's semantic interest. Moreover, the subject located at lower bound levels has more accurate semantic specificity description ability than that at upper bound levels. Hence, the semantic specificity ability of a lower bound subject with more sibling nodes is large than that of an upper bound subject with less brother nodes.

Algorithm 1: Analyzing semantic coverage degree of subject.

Input: ontology $\Theta = (S, R, A)$, a coefficient $\lambda_1 > \lambda_2 > 1$.

Output: $Scd(S)$ applied to ontology.

1. Set $k=1$, get the root set S_0 of S from Θ , for $s \in S_0$, set $Scd(s) = k$.
2. Remove S_0 , get the new root nodes set S' from Θ
3. **If** $S' = \emptyset$ then **return**;
4. **For each** $s' \in S'$ **do**
 Get the parent node s_0 of s' and sibling nodes set Ss_0 .
If $s' \xrightarrow{Is-a} s_0$,
 set $Scd(s') = \lambda_1 \times Scd(s_0) \times \log(1 + |Ss_0|)$
If $s' \xrightarrow{Part-of} s_0$,
 set $Scd(s') = \lambda_2 \times Scd(s_0) \times \log(1 + |Ss_0|)$
5. $S_0 = S_0 \cup S'$, go to step 2.

In Algorithm 1 above, the semantic specificity degree of target subject inherits from the parent node with different propagation strengths according to different levels and different number of siblings. However, semantic interest is reflected by semantic strength of subject relying on user's behaviors. Semantic strength degree is the accumulation of semantic specificity degree. Based on this idea, taking into account occurrence frequency of subject, semantic strength degree of a subject for user u on subject s is measured as:

$$Ssd_s(u) = \sum_{m \in D_s^u} Scd(s) \times \eta(s, m) \quad (6)$$

where $\eta(s, m) = 1$ if $s \in m$; otherwise $\eta(s, m) = 0$.

b) Semantic Interest Degree of Subject

Taking into account all subjects user involved in, semantics interest degree of subject s for user u can be shown in Eq. (7):

$$Sid_s(u) = \frac{Ssd_s(u)}{\max_{s' \in S_u} \{Ssd_{s'}(u)\}} \quad (7)$$

where S_u is the set of subjects user u are interested in.

According to the analysis above, we can represent ontology user model of a user in terms of content interest and semantic interest for some subjects.

B. Closeness Metric of User Model

1) Closeness of Content Interest

In this subsection, a similarity metric strategy about content interest is used to judge users' closeness for some subject. For a specific subject, as two user models are both interested in it, the similarity between models is supposed to be large. Moreover, when the interest degrees of user models are both consistently strong, the users have more common interests. Hence, the similarity metric can be described as Eq. (8):

$$Csim_s(u_i, u_j) = 1 - \frac{|Cid_s(u_i) - Cid_s(u_j)|}{\max\{Cid_s(u_i), Cid_s(u_j)\}} \quad (8)$$

2) Closeness of Semantic Interest

For the purpose of evaluating the similarity of ontology interest tree, we not only consider semantic interest degree of subject in the interest tree, but also pay close attention to the structure of interest tree. Different interest structures play an important role in different similarity calculation. The structure of children nodes can reflect semantic affinity levels among user models. A user model is different from another one on some subjects with their children nodes different. For example, although two users are both interested in "sports", one likes "basketball" while the other prefers "football", the interest similarity of two users is close only from perspective of "sports" subject, which even is viewed as 1. Actually, there is a huge difference between each other if interest tree structure is taken into account. Hence, similarity of interest structure is an important factor for describing closeness between user models.

a) Self-Node Similarity of Subject

If both of users are interested in the subject s , the self-structure similarity of subject is 1, which is shown as in Eq. (9):

$$Snsim_s^1(u_i, u_j) = \begin{cases} 1 & \text{if } Cid_s(u_i) > 0 \text{ and } Cid_s(u_j) > 0 \\ 0 & \text{else} \end{cases} \quad (9)$$

b) Children Nodes Similarity of Subject

The children nodes of subject that user is interested in represent the degree the user exert all one's energies for target subject. In this research, firstly, we consider two users have intimate relationship if they both have the same interest subjects; secondly, the structure of parent subject depends on those of children nodes severely. The process of structure similarity between user models is aggregating for total intersection subjects in each ontology layer, as in Eq. (10):

$$Snsim_s^2(u_i, u_j) = \frac{\sum_{k \in Layer} \frac{|S_i^k \cap S_j^k| \cdot k}{|S_i^k \cup S_j^k|}}{|Layer|} \quad (10)$$

where $S_i^k = \{s' \mid s' \rightarrow \dots \rightarrow s, Cid_s(u_i) > 0\}$, which is the set of subjects user u_i is interested in and all of them are children nodes of subject s in layer k ;

$S_j^k = \{s' \mid s' \rightarrow \dots \rightarrow s, Cid_s(u_j) > 0\}$ is the set of subjects user u_j is interested in and all of them are children nodes of subject s in layer k ; and

$Layer = \{n \mid \exists s' \rightarrow \dots \rightarrow s, Cid_s(u_i) > 0 \text{ or } Cid_s(u_j) > 0\}$ is the set of layer n in which subject s' arrive in subject s by n steps. l is the layer number of systematic original ontology tree. $|\bullet|$ is the cardinal number of the set \bullet .

Finally, the node structure similarity degree is measured by the weight sum of self-node and children nodes similarity degree, as is in Eq. (11):

$$SNSim_s(u_i, u_j) = \alpha Snsim_s^1(u_i, u_j) + (1 - \alpha) Snsim_s^2(u_i, u_j) \quad (11)$$

c) Semantic Similarity of Subject

The semantic similarity of subject is calculated by combining the similarity of semantic strength between two user models with that of node structure. Hence, the semantic similarity of subject between u_i and u_j is shown as:

$$Ssim_s(u_i, u_j) = SSsim_s(u_i, u_j) \cdot SNSim_s(u_i, u_j) \quad (12)$$

where the semantic strength similarity can be represented as:

$$SSsim_s(u_i, u_j) = 1 - \frac{|Sid_s(u_i) - Sid_s(u_j)|}{\max\{Sid_s(u_i), Sid_s(u_j)\}} \quad (13)$$

The semantic similarity of subject reflects user's semantic connection relationships from perspective of entire interest

tree structure. As is described in Eq. (12), the bigger semantic strength between each other and the greater interest structure between each other, the larger semantic similarity will be.

3) Multi-granularity Similarity

To calculate the similarity between user models for a specific subject, the content similarity and semantic similarity need to be further aggregated with appropriate weighting distribution. When one of users has some strong interest for subject s , the closeness degree of attention content is an important factor to measure their intimacy. Instead, when both of users are indifferent to the target subject, the semantic similarity will be relative important to determine their proximity. Therefore, the weighting parameter combination is shown as Eq. (14):

$$Sim_s(u_i, u_j) = \beta Csim_s(u_i, u_j) + (1 - \beta) Ssim_s(u_i, u_j) \quad (14)$$

where $\beta = \max\{Cid_s(u_i), Cid_s(u_j)\}$, and parameter β represents the relative importance of two factors.

Similarity between user models should be reflected by multi-granularity similarity deriving from multi-granularity subjects. Taking into account mutual shared subjects by both user models, the similarity of each subject is calculated to supply similar friends about every subject.

V. EXPERIMENTS

A. Experiment Data

Our experiments use Sina Weibo data to recommend friends relating to sports domain. Firstly, In October 2012, the data about almost 1900 users was used to gather extended data from Sina platform using API, which involved in user's friends, user's fans and user's reposting micro-blogs. However, some spammers and bots users are common phenomena in micro-blogs, so we select 4175 appropriate users in whom 98 users have friends and fan both more than 40 to reduce noise and bias in our experiment. Secondly, 60 thousand micro-blogs about above users were preprocessed by some steps such as removing stopwords, special symbols (*, #, @, /, &, etc.), digits and so on to extract interest subjects for conducting 4175 user profiles. Lastly, in specific sports domain, 98 user profiles were used to verify performance of multi-granularity friend recommendation.

B. Experiment Design

The proposed multi-granularity similarity method is used to find the most similar top k users to recommend the target user profile. In conventional recommender systems, a common method for top k users were recommended by similarity of unify granularity subject. For verifying the performance of recommendation friends' quality, some experimental comparisons against conventional method are conducted. In our research, top k similar users of user u are deriving from children subjects set $Chid(s)$ of target subject s . For each sub-subject, the number of users recommended can be denoted as k_i in Eq. (15):

$$k_i = \frac{Cid(s_i)}{\sum_{s_i \in Chid(s)} Cid(s_i)} \times k \quad (15)$$

where $Chid(s) = \{s' \mid s' \rightarrow s, Cid(s') > 0\}$ and $\sum k_i = k$.

In the proving experiment, we use precision and recall to evaluate recommendation accuracy. These metrics have been defined as follows:

$$Precision = \frac{|F_A \cap F_R|}{|F_R|} \quad (16)$$

$$Recall = \frac{|F_A \cap F_R|}{|F_A|} \quad (17)$$

where F_R is the recommended friends set and F_A is the actual friends set for each user. Here, the members of friends set are sum of followee users.

C. Experiment Results and Evaluations

To test the accuracy of recommendation similar friends, we make experiments on followee users set of target user. In our paper, we performed an experiment by varying the value of recommended user list size k from 3 to 15 with an increment of 3 in order to find appropriate number of recommended users. Figure 3 presents average precision curves of 98 user models at different k in different granularity subjects of sports. Especially, sports subject covers a lot of specific subjects, average precision based on multi-granularity subject is prior than traditional a single subject.

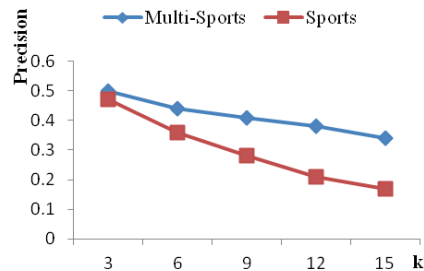


Figure 3. Average precision curves of 98 user models at different k values.

Correspondingly, Figure 4 show the average recall curves. As we can see in Figure 4, recall value is gradually increasing with the increment of value k , which states that inverse consistence with precision curves. Although the number of actual followee friends is larger for each user profile, we select 15 friends for every user randomly to conduct our experiment. In Figure 4, the result of multi-

granularity recommendation method is more successful than single subject recommendation in the limited users dataset.

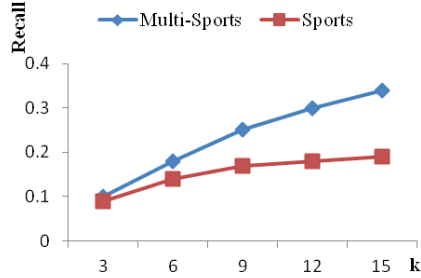


Figure 4. Average recall curves of 98 user models at different k values.

VI. CONCLUSION

In this paper, a multi-granularity recommendation mechanism based on multi-granularity similarity of subjects among user models is proposed to improve the satisfactory degree of end-user in personalized services. The proposed method firstly conducts user model by combing content interest and semantic interest. A semantic coverage degree method is introduced for semantic interest discovery. For closeness of subject similarity measuring, interest structure of user model is compared to detect similarity of semantic space. Top k users in multi-granularity similarity of subjects is recommended, which is compared against the recommendation of single subject in performance of precision and recall measure. The experiment results demonstrate that our proposed recommendation mechanism is promising.

In future work, we will investigate these methods by introducing user model's social friend factor to fix the role of interest subject in the recommendation service. The social friend relationships is useful to implement modification and optimization of recommendation results. Hence, the investigation will extend the application of the proposed methods and increase the contribution and significance of the present work by considering the social relationships of users.

REFERENCES

- [1] Y. Li and N. Zhong, "Mining Ontology for Automatically Acquiring Web User Information Needs," *IEEE Trans. Knowledge and Data Engineering*, 2006, vol. 18, no. 4, pp. 554-568.
- [2] Y. Li and N. Zhong, "Web Mining Model and Its Applications for Information Gathering," *Knowledge-Based Systems*, 2004, vol. 17, pp. 207-217.
- [3] Hak-Lae Kim, John G. Breslin, Stefan Decker, and Hong-Gee Kim. Mining and Representing User Interests: The Case of Tagging Practices, *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 2011, vol. 41, no. 4, pp. 683-692.
- [4] Xiaohui Tao, Yuefeng Li, and Ning Zhong. A Personalized Ontology Model for Web Information Gathering. *IEEE Trans. Knowledge and Data Engineering*, 2011, vol. 23, no. 4, pp. 496-511.
- [5] A. Sieg, B. Mobasher, and R. Burke, "Web Search Personalization with Ontological User Profiles," *Proc. 16th ACM Conf. Information and Knowledge Management (CIKM '07)*, pp. 525-534, 2007.
- [6] Leung, K.W.-T. and Dik Lun Lee, "Deriving Concept-Based User Profiles from Search Engine Logs," *IEEE Trans. Knowledge and Data Engineering*, 2010, vol. 22, no. 7, pp. 969-982.
- [7] Elias Iosif, Alexandros Potamianos, Unsupervised Semantic Similarity Computation between Terms Using Web Documents. *IEEE Trans. Knowledge and Data Engineering*, 2010, vol. 22, no. 11, pp. 1637-1647.
- [8] K.-C. Siu and H. Meng, "Semi-Automatic Acquisition of Domain-Specific Semantic Structures," *Proc. European Conf. Speech Comm. and Technology*, 1999, pp. 2039-2042.
- [9] A. Pargellis, E. Fosler-Lussier, C.-H. Lee, A. Potamianos, and A. Tsai, "Auto-Induced Semantic Classes," *Speech Comm.* 2004, vol. 43, no. 3, pp. 183-203.
- [10] D. Lewis, "Naive Bayes at Forty: The Independence Assumption in Information Retrieval," *Proc. European Conf. Machine Learning*, 1998, pp. 4-15.
- [11] Stephen J. Green. Building Hypertext Links By Computing Semantic Similarity. *IEEE Trans. Knowledge and Data Engineering*, 1999, vol. 11, no. 5, pp. 713-730.
- [12] Yin Guisheng, Sheng Qiuyan. Research on Ontology-Based Measuring Semantic Similarity. 2008 International Conference on Internet Computing in Science and Engineering, 2008, pp. 250-253.
- [13] Min Liu, Weiming Shen, Qi Hao, Junwei Yan. An weighted ontology-based semantic similarity algorithm for web service. *Expert Systems with Applications*. 2009, vol. 36, pp. 12480-12490.
- [14] P. Massa, B. Bhattacharjee, Using trust in recommender systems: an experimental analysis, *iTrust 2004*, 2004, pp. 221-235.
- [15] A. Albadvi, M. Shahbazi, A hybrid recommendation technique based on product category attributes, *Expert Systems with Applications*. 2009, vol. 36, no. 9, pp. 11480-11488.
- [16] W.S. Yang, J.B. Dia, Discovering cohesive subgroups from social networks for targeted advertising, *Expert Systems with Applications*. 2008, vol. 34, no. 3, pp. 2029-2038.
- [17] S. Debnath, N. Ganguly, P. Mitra, Feature weighting in content based recommendation system using social network analysis, *Proceeding of the 17th International Conference on World Wide Web*, 2008, pp. 1041-1042.
- [18] D.-R. Liu, P.-Y. Tsai, P.-H. Chiu. Personalized recommendation of popular blog articles for mobile applications. *Information Sciences*, 2011, vol. 181, pp. 1552-1572.
- [19] Kebo Mei, Bofeng Zhang, Jianxing Zheng, et al. Method of Recommend Microblogging Based on User Model. 2012 IEEE 12th International Conference on Computer and Information Technology (CIT), 2012, 27-29 Oct, pp.1056-1060.
- [20] Luxu Zhang, Bofeng Zhang, Jianxing Zheng, et al. Discovering Similar User Models Based on Interest Tree. 2012 IEEE 12th International Conference on Computer and Information Technology (CIT), 2012, 27-29 Oct, pp. 1046-1050.