Proceedings of the 33rd Chinese Control Conference
July 28-30, 2014, Nanjing, China

A Hybrid Recommendation Approach for Network Teaching Resources Based on Knowledge-Tree

ZHANG Haidong[1], NI Wancheng[1], ZHAO Meijing[1], LIU Yu[2], YANG Yiping[2]
1,CASIA-HHT Joint Laboratory of Smart Education
2, Integrated Information Research Center, Institute of Automation Chinese Academy of Science, Beijing, 100190, China
{ haidong.zhang, wancheng.ni, meijing.zhao, yu.liu, yiping.yang}@ia.ac.cn

Abstract: Recommender systems could be used to help learners or teachers find useful network teaching resources effectively in technology enhanced learning (TEL), but the quality of recommendations is always limited by cold start, data sparsity, lack of learning or teaching contextual aware, and so on. Considering the features of network teaching resources, a hybrid recommendation approach is presented in this paper. The presented approach takes user context, association rules between resources, association rules between resources and the structure of lessons into consideration, and is mainly composed of five modules. These five modules are: (1) Course model, which is used to express the structure of lessons; (2) Association rules between resources, which are discovered in resources association rule mining module; (3) Association rules between resources and lessons, which are discovered in lessons and resources association rule mining module; (4) User dynamic profile, namely, user context which are found in reasoning user dynamic profile module; (5) Hybrid recommendation, which generates recommended lists in hybrid recommendation module. Finally, experiments have been done on a real dataset from "HHT Education Cloud", an enterprise education resources sharing platform. The results have shown that our hybrid method can outperform the general recommendation method. The Recall has an improvement ranging from 0.131 to 0.213, and the Precision has an improvement ranging from 0 to 0.152, when the number of recommendations changes from 1 to 40.

Key Words: Recommender Systems, Knowledge-Tree, Dynamic Profile, Association Rule Mining

1. Introduction

The rapid development of information technology and increase of web sites brings tremendous information overload. Getting appropriate information on demand is becoming more and more difficult. It is no doubt that the emerging recommender systems are the effective approaches to overcome the information overload. Hence, recommender systems get rapid development, and have been largely used in many areas nowadays to generate interesting items for every user. For instance, e-commercial companies personalize the online stores to provide customers with their products (such as books, cameras, computers etc.) [1], entertainment companies recommend movies, music, and television programs to users, some websites provide personalized newspapers, blogs, community services, friends recommendation and so on [2]. There are mainly four prominent recommendation approaches nowadays [3]:

- Content-based recommendation approaches: the active user is recommended items similar to his/her preferred ones;
- Collaborative recommendation approaches: the active user is recommended items that people with similar tastes or preferences liked in the past;
- Knowledge recommendation approaches: the active user is recommended items based on some additional knowledge and constraint conditions exploited from his/her historical behavior.
- Hybrid approaches: these approaches combine the above any two or three methods in different manners.

As recommender systems become more and more popular, research interests are aroused in technology enhanced learning (TEL) which covers technologies that support and enhance learning practices and teaching activities [4], such as multimedia learning, online education, computer-assisted instruction, etc. TEL presents personal services with the users' preferences more accurately and intelligently, integrated into recommender systems.

The application of recommender systems for technology enhanced learning (TEL) can help learners find network resources they are interested in. However, it is difficult to simply transfer a recommender system from the commercial area to TEL. This is because several particularities and complexities exist in TEL, resulting in meeting the needs of the target users with difficulties. They can be abstracted into three aspects in TEL recommendation. They are as follows:

- There are strongly dependencies on user context and other conditions or rules of learning in TEL. For example, teachers of different subjects have different teaching strategies.
- Learning is a continuous process of knowledge accumulation. The interests of users and the content of learning objects always change over time along with users' improving levels. It is unwise to recommend resources of knowledge mastered by learners. But such problems do not exist in commercial recommendation.
- The attributes of networked resources such as subject, author and type should be considered in order to improve the recommendation quality in TEL.

In formal TEL, a teacher has to fulfill some certain educational goals to prepare a lesson. He or she need not only find related content to illustrate the lesson, but also visualize and present some information to get learners across the lesson. The user context and associations between lessons and resources may not be got by the above four approaches, which limits the quality of recommendations.

Hence, a recommender approach is needed to recognize user context and mine the associations between lessons and teaching resources. On the one hand, the results generated by mining the associations between lessons and teaching resources could filter the massive teaching resources effectively according to user context. On the other hand, the associations between user context and lessons are acquired to predict future prepared lessons. Meanwhile the approach is also needed to extend the recommendations with the increase of users' behavior records.

The paper is organized as follows: Section 2 presents some previous research related to recommender systems in TEL. Section 3 describes the architecture of our recommender system and recommendation approach. Section 4 describes some experimental tests and evaluates the recommendation approach's effectiveness. Finally, section 5 outlines some conclusions and further research.

2. Related Works

Many research groups have integrated ontology concepts into recommender systems for TEL. Aroyo L. et al. presented an AIMS model that consists of domain model, course sequencing model and resource management model with ontological structures representing knowledge. The ontology concept was applied to construct the Authoring Task Ontology (ATO) and achieve flexibility in reasoning and higher usability of the authoring environments for Intelligent Educational Systems [5]. Daniel Lemire et al. proposed a learning resources personalized recommendation system to improve the teaching quality of E-Learning System. The system combined p2p technology and ontology semantic description to improve service quality [6]. Ontology-based recommender methods took user context and specific domain into account, which improved the quality of recommendations effectively even with little user information available. However, designing and implementing ontology are often done manually by experts and require professional knowledge. Meanwhile, high maintain effort, reengineer and adaptation to user and domain preferences also bring difficulties to recommender system, especially for recommendation in large educational data sets.

On the other hand, some generic data mining methods are combined with the main recommender ones to provide users with recommender lists. Daniel Lemire, et al. developed RACOFI (Rule-Applying Collaborative Filtering) which was composed of collaborative filtering algorithm and RALOCA (Rule Applying Learning Object Comparison Object) to generate context-aware recommendation lists. This algorithm helped users discover what they were looking for through explicit ratings and inference rules [7]. Mei-Hua Hsu showed an online personalized English learning recommender system to provide learners with reading lessons and increase the motivation to learn. This system analyzed learners' reading data to find learners' interests by combining content-based approach, collaborative filtering algorithm, and data mining techniques (clustering and association rules algorithm) [8]. Majda Maatallah and Hsssina Seridi-Bouchelaghem proposed a recommendation method which combines a fuzzy-based CF algorithm and CB one to make better recommendations. It recommended items with multi-context according to learners' preferences and importance of knowledge [9]. A new recommendation framework was proposed by analyzing web log data. It integrated k-nearest neighbors, Markov model and association rule mining to make accurate recommendations [10]. Compared with ontology-based recommender systems, ones combined with data mining methods could be used for large education data sets without manual maintain. But some recommendations generated by the recommender systems may be outdated because of the ignorance of user context. For example, if one knowledge point has been mastered by user, it is unadvisable to recommend learning materials related to this point.

Meanwhile, considering the features of formal TEL, lessons are organized in sequence based on pre-defined sequence in formal TEL. For example, the lesson *Unit 2 What's the matter?* is ahead of the lesson *Unit 3 What are you doing for vacation?* in English book (PEP edition) in eighth grade. Teachers usually prepare their lectures in accordance with the sequence. Hence, teaching resources should be recommended to teachers in this sequence. However, registered users' profiles are always incomplete or inaccurate. That restricts some methods achieving users' interests or tracking users' levels by analyzing users' log data. Hence, we propose a hybrid approach consists of content-based method and data mining one based on knowledge-tree, which is different from the aforementioned combination methods. The innovations of this approach are listed as follows: a new course model (knowledge-tree) is proposed, the association rules between lessons and resources are mined, user dynamic profile is reasoned to predict user's interests, and the association rules are mined between resources to extend recommendation.

3. Our Recommendation System

The overall system architecture is shown as Fig. 1. The hybrid method could filter teaching resources in terms of user context, and extend the recommendations by mining the association rules between resources. This method is mainly based on seven modules: (1) Course model, which is used to express content features of lessons, is organized with knowledge tree organizing the structure of lessons and lesson tags expressing the content features. (2) Resource model is expressed as content features extracting from some description about resources. (3) Lessons and resources association rule mining is used to mine the relation between lessons and resources. (4) User profiles consist of static profile filled in by users, and dynamic profile used to express user context. (5) Reasoning user dynamic profile could identify user context to find suitable resources effectively. (6) Resources association rule mining, which discovers association rules between resources, extends the recommendations. (7) Hybrid recommendation generates the final recommendations by combining the above modules with mixed and weighted approaches.

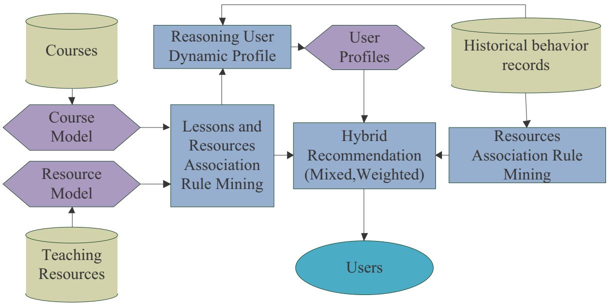


Fig. 1: Hybrid recommender method

3.1 Course Model and Resource Model

As described in section 2, recommending teaching resources should be in the sequence of lessons. Hence, course model was proposed, including knowledge-tree indicating the structure of textbooks, and lesson tags indicating the features of every lesson in textbooks, as Fig. 2. Knowledge-tree is organized with a decision tree. In this tree, each leaf node is assigned a lesson label. The non-terminal nodes, which include the root and other internal nodes, contain attribute test conditions to separate lessons that have different characteristics. These characteristics consist of grade, subject, version, unit, and lesson. Every leaf node connects with a tags set in lesson tags. The tags come from the analysis of webpages about corresponding lessons.

Resource model consists of user's descriptions and resource tags extracted from user's descriptions. Lesson tags and resource tags are used for calculating the association degree between lessons and resources.

3.2 Lessons and Resources Association Rule Mining

In formal teaching process, a teacher prepares one certain lesson, and searches teaching resources about this lesson. Hence, we can imitate search engine (SE) to find these teaching resources and filter them by other methods so as to achieve results better matching users' interests. We define the form L→R as an association rule, where L is one path of knowledge-tree and R is one of teaching resources. The association rule can be measured in terms of the following formula:

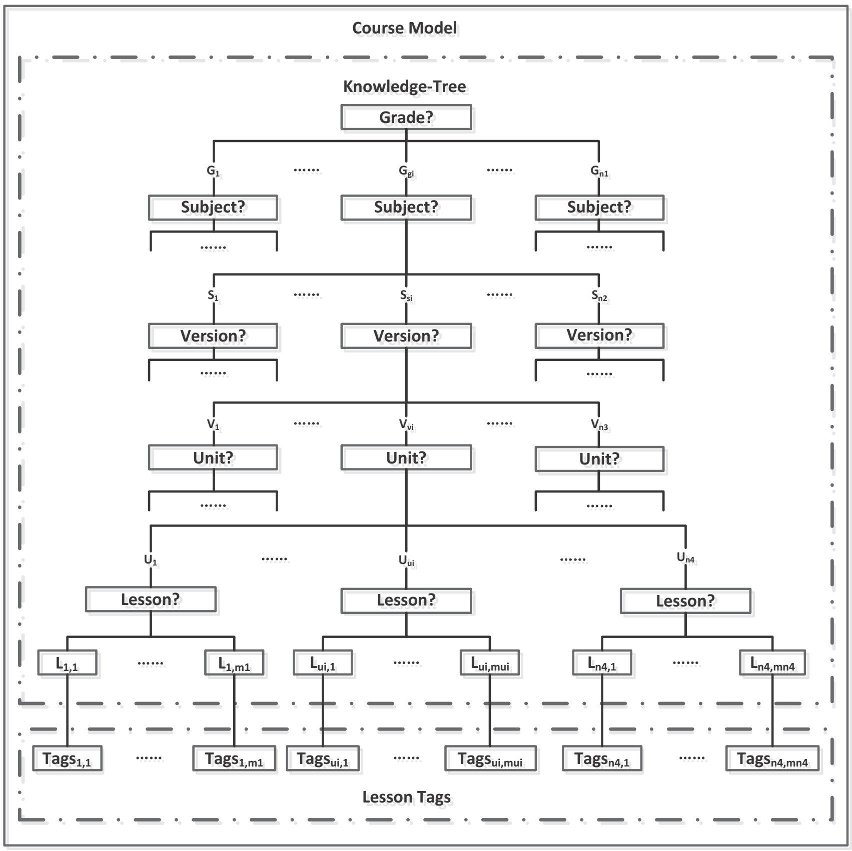


Fig. 2: Course Model

$$\mathrm{LRAW}_{\mathrm{L,R}} = \alpha \times \frac{Num(\mathrm{Title}_L \cap Tags_R)}{Num(\mathrm{Title}_L)} + (1-\alpha) \times \frac{Num(\mathrm{Tags}_L \cap Tags_R)}{Num(Tags_R)} \quad (1)$$

Where $\mathrm{LRAW}_{\mathrm{L,R}}$ denotes the association between lesson L and resource R. $\mathrm{Title}_\mathrm{L}$ denotes the title tags set of lesson L. Tags_R denotes the tags set of resource R. Tags_L represents the tags set of lesson L.α is constant. Num(*) represents the number of elements in a set. The generated rules express the resources list of every lesson. At last, each lesson can be associated with a set of teaching resources.

$$RS_L = \{R_j : \mathrm{LRAW}_{\mathrm{L,j}} \mid \mathrm{R_j} \in \mathrm{Re}\,sources, LRAW_{L,j} > \mathrm{minValueL}\} \quad (2)$$

Where RS_L represents the resources set of lesson L, R_j represents one of the resources associated. Resources represents the set of all teaching resources. minValueL represents a minimum weight threshold.

3.3 User Profiles and Reasoning User Dynamic Profile

User profiles include a static profile and a dynamic profile. A static profile includes a user's name, gender, address, etc. They are filled in by users. A dynamic profile is one path in the knowledge-tree, expressing lessons a user is concerning on.

The user dynamic profile is reasoned through analyzing his or her behavior records. Our algorithm is defined as follows:

Step 1: Extract every user's behavior records, and represent them with a vector.

$$UD_K = (\mathrm{R}_1, \mathrm{R}_2, ..., \mathrm{R}_j, ..., \mathrm{R}_N) \quad (3)$$

Where UD_K represents the vector of the resources user K has historical behaviors on, and R_j represents one of the resources.

Step 2: Knowledge-tree is organized according to lessons in each textbook. Hence, each textbook could be indicated as the following vector.

$$GSV_T = (L_1, L_2, ..., \mathrm{L}_j, ..., L_N) \quad (4)$$

Where GSV_T denotes the lessons vector of the textbook T, and L_j denotes the jth lesson in textbook T. The attribute of textbook T includes grade, subject and version. All textbooks could be represents as formula (4).

$$GSVS = \{\mathrm{GSV}_T \mid \mathrm{T} \in \mathrm{textbooks}\} \quad (5)$$

Where GSVS denotes the set of all textbooks, and T denotes one of the textbooks.

Step 3: Get the vector of resources about lessons GSV_j from section 3.2.

$$\mathrm{R}V_{\mathrm{GSV}_T} = (\mathrm{RS}_{L_1}, \mathrm{RS}_{L_2}, ..., \mathrm{RS}_{\mathrm{L}j}, ..., \mathrm{RS}_{L_N}) \quad (6)$$

Where $\mathrm{R}V_{\mathrm{GSV}_T}$ denotes the resources set vector of the GSV_T and RS_{Lj} generated by formula (2) denotes the resources set of lesson L_j.

Step 4: Calculate the intersection of formula (2) and formula (6) with the following formula:

$$URS_{\mathrm{K,T}} = set(UD_K) \cap set(\mathrm{R}V_{\mathrm{GSV}_T}) \quad (7)$$

$$URS_K = \{\mathrm{URS}_{\mathrm{K,T}} \mid \mathrm{GSV}_T \in GSVS\} \quad (8)$$

Where $URS_{K,T}$ denotes the intersection of user K resources set and the resources set of the GSV_T. URS_K denotes the set of $URS_{K,T}$. We can find the most interested textbooks of user K based on the number of elements in $URS_{K,T}$ of URS_K. Then the grade, subject and version of the user dynamic profile can be reasoned as GSV_K based on the attribute of textbooks.

Step 5: Combined with the the interested textbook GSV_K, the intersection of formula (2) and RS_{Lj} in $\mathrm{R}V_{\mathrm{GSV}_K}$ is calculated with formula (9).

$$URLS_{K,j} = set(UD_K) \cap set(RS_{Lj}) \quad (9)$$

Then we can find the lessons the user taught in the past, the lesson the user is teaching and the lessons the user will teach in GSV_K. At last, user K dynamic profile is reasoned by GSV_K and the lesson user K is teaching as $GSVL_K$.

3.4 Resources Association Rule Mining

When users are preparing for a certain course, they will search lots of web teaching resources in a short time to enrich lessons and get lessons across easily. These resources center on a common topic. Hence, we assume they are associated, that is to say, the resources one user browses, rates or downloads in a short time are regarded as one transaction which is similar to market basket transactions commonly known in data mining. Every user's behavior records can be divided into some transactions. Then, we define the form $X \leftrightarrow Y$ as an association rule, where X and Y are two different teaching resources. The association rule can be measured in terms of its associated weight which determines how frequently resource X appears in transactions that contain resource Y or otherwise. The formal definition of associated weight is computed as follows:

$$RRAW_{X,Y} = \frac{Num(\mathrm{T}_{X \cap Y})}{Num(\mathrm{T}_{X \cup Y})} \quad (10)$$

where $RRAW_{X,Y}$ is the associated weight between resource X and resource Y. $Num(\mathrm{T}_{X \cap Y})$ is the number of transactions which both resource X and resource Y appear in. $Num(\mathrm{T}_{X \cup Y})$ is the number of transactions which either resource X or resource Y appears in.

$$Num(\mathrm{T}_{X \cup Y}) = Num(T_X) + Num(\mathrm{T}_Y) - Num(\mathrm{T}_{X \cap Y}) \quad (11)$$

where $Num(\mathrm{T}_X)$ is the number of transactions which resource X appears in, $Num(T_Y)$ is number of transactions which resource Y appears in.

3.5 Hybrid Recommendation

Hybrid recommender method is based on user profiles, the association between lessons and resources and the association between resources. The historical behavior records of user K are expressed as a vector formula (3).

Then we can recommend user K with network teaching resources. The algorithm is as follows:

Step 1: According to section 3.3, user K dynamic profile can be reasoned as $GSVL_K$. Then which lesson is taught and which lessons will be taught in the near future could be found. Combining with the association between lessons and resources, choose the resources associated with user dynamic profile which is one path of knowledge-tree. The initial recommending resources can be expressed as the following set.

$$RL_K = \{\mathrm{R}_i : LRAW_{\mathrm{K,i}} \mid R_i \in \mathrm{Resources}, \mathrm{LRAW}_{\mathrm{K,i}} > \mathrm{minValueL}\} \quad (12)$$

Where RL_K denotes recommendation lists based on user K dynamic profile, and $LRAW_{K,i}$ denotes the association weight between user dynamic profile and resource R_i. minValueL is a minimum weight threshold.

Step 2: According to the association between resources, calculate the resources associated with R_j in formula (3). Then we can get the associated weight.

$$RA_K = \{\mathrm{R}_{j,k} : \mathrm{RRAW}_{\mathrm{j,k}} \mid \mathrm{R}_{\mathrm{j,k}} \in \mathrm{Resources}, \mathrm{RRAW}_{\mathrm{j,k}} > \mathrm{minValueR}\} \quad (13)$$

Where RA_K denotes recommendation lists based on the association between resources, $R_{j,k}$ denotes one of elements associated with resource R_j, and $RRAW_{j,k}$ denotes the associated weight based on section 3.4 between resource R_j and resource R_k. Resources represents the set of all teaching resources. minValueR is a minimum weight threshold.

Step 3: Combine formula (12) and formula (13) with weighted approach to get a new recommendation list.

$$Rec_K = (1-\beta)\times RL_K + \beta \times RA_K \tag{14}$$

Step 4: We can get the final recommendation list by Inversed File Retrieval Algorithm based on the weights in formula (14).

4. Experiments and Evaluation

A real world dataset is applied in our experiments. The dataset of behavior records comes from "HHT Education Cloud", an enterprise education resources sharing platform[3]. It contains 812359 behavior records. In our experiment, 90% of the dataset is used for training and 10% for testing.

The evaluations of recommendation performances are consisted of Precision and Recall. The Precision and Recall are defined as follows:

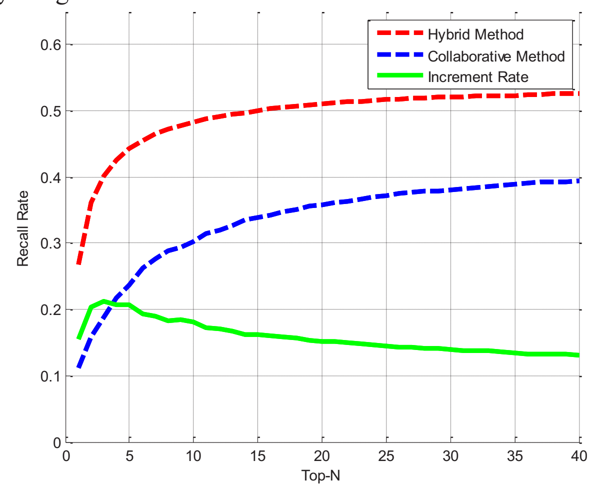
$$Precison = \frac{\sum_u R(u)\cap T(u)}{\sum_u R(u)}$$

$$Recall = \frac{\sum_u R(u)\cap T(u)}{\sum_u T(u)}$$

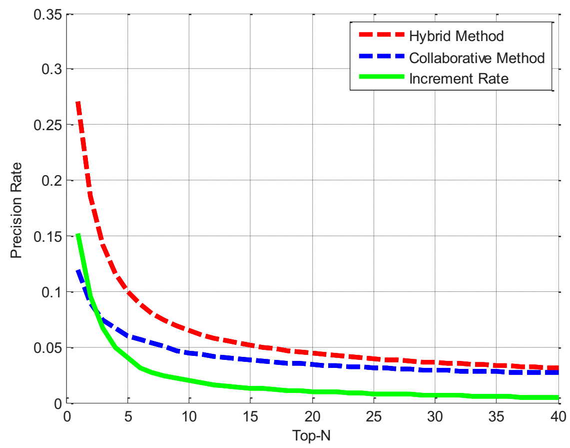
We compare the recommendation results between our hybrid method and a general collaborative method (user-based collaborative method). Because of the data sparsity, some recommendations may be really needed but lack of behavior records for users, resulting in the precision value on the low side. Hence, the parameters optimized are $\alpha = 0.5$, $\beta = 0.7$, minValueL = 0.3 and minValueR = 0.5 through training based on the Recall. Fig. 3 reports the recall values and precision values at top-N resources. The results clearly show the advantages of our hybrid method in recall rate and precision rate.

The Recall goes up with the increased number of recommendations by either general method or our hybrid method. The difference between the hybrid method and the general method represented by a green solid line in fig. 3 (a) ranges from 0.131 to 0.213. As mentioned above, the data sparsity has a negative effect on the Precision. In fig. 3 (b), the Precision declines with the increased number of recommendations by either general method or our hybrid method. But the Precision generated by our method in general has improved, and the increment range is 0 to 0.152 which is described by a green solid line in fig. 3 (b).

There are two reasons to explain these improvements. On the one hand, the process of seeking k-nearest neighbor in the general method is always adversely affected by the data sparsity. While the resources association rule mining module in the hybrid method transfers the historical records to market basket transactions commonly known in data mining. The relation between resources could be discovered by mining frequent-2 itemsets, which alleviates the sparsity problem to some extent. On the other hand, user context could be detected by reasoning user dynamic profile module. We can thereby predict which lessons will be prepared for users based on knowledge-tree, and recommend relevant resources. It cannot be accomplished by the general method.



(a) Recall Rate



(b) Precision Rate

Fig. 3: Hybrid Method and Collaborative Method

5. Conclusions

Along with the development of recommender systems in many areas, many solutions to deal with the recommender problems in technology enhanced learning have been proposed. Considering the structure of lessons and the network teaching resources information, we propose a hybrid method based on the knowledge-tree and user dynamic profiles. It also combines content-based recommendation method with data mining. This method alleviates the sparsity problem and can predict users' goals in the near future to some extent. The experiments on real datasets show that our algorithm can be better than the traditional recommendation algorithms in precision and recall.

Analyzing our system, we could conclude that the recommendation results are limited mainly by the quality of lesson tags and resource tags which are used to mine the association between lessons and resources. With regard to future research, we believe that natural language understanding could effectively extract meaningful tags and filter function words to mark lessons or resources completely or accurately.

[3] http://www.910edu.com/

Meanwhile, some new recommendation methods will be explored to improve the quality of recommendations in TEL.

Acknowledgements

We would like to thank Hong He Technology for providing the dataset that supported our research work.

References

[1] Sanjeev Kumar Sharma, Dr. Ugrasen Suman, Design and Implementation of Architectural Framework of Recommender System for e-Commerce, *International Journal of Computer Science and Information Technology & Security(IJCSITS)*, 1(2): 153-162, December 2011.

[2] Jiahui Liu, Peter Dolan, Elin Ronby Pedersen. Personalized News Recommendation Based on Click Behavior. *IUI 10 Proceedings of the 15th international conference on Intelligent user interfaces*: 31-40, 2010.

[3] Dietmar Jannach, Markus Zanker, Alexander Felfernig, Gerhard Friedrich, *Recommender Systems An Introduction.* First edition: Cambridge University Press, Chapter 1, 2011.

[4] Francesco Ricci, Lior Rokach, Bracha Shapira, Paul B.Kantor, *Recommender Systems Handbook*, Springer Science Business Media: 392, 2011.

[5] Aroyo L., Mizoguchi R., Tzolov, C, OntoAIMS: Ontological Approach to Courseware Authoring, I*nternational Conference on Computers in Education 2003,* December 2003.

[6] Qing Yang, Yuan Yuan, Junli Sun, KaiMin Cai, Semantic P2P-based Learning Resources Personalized Recommendation System Design, *The Third Pacific-Asia Conference on Circuits, Communications and System (PACCS) 2011*: 1-4, July 2011.

[7] Daniel Lemire, Harold Boley, Sean McGrath, Marcel Ball, Collaborative Filtering and Inference Rules for Context-Aware Learning Object Recommendation, *International Journal of Interactive Technology and Smart Education*, 2(3), August 2005.

[8] Mei-Hua Hsu, A personalized English learning recommender system for ESL students, E*xpert Systems with Applications:* 683–688, October 2006.

[9] Majda Maatallah, Hsssina Seridi-Bouchelaghem, Enhanced Collaborative Filtering to Recommender Systems of Technology Enhanced Learning, *The 4th International Conference on Web and Information Technologies 2012*: 129-138, April 2012.

[10] Anitha, N.Krishnan, A Web Usage Mining based Recommendation Model for Learning Management Systems, *IEEE International Conference on Computational Intelligence and Computing Research (ICCIC) 2010*: 1-4, December 2010.