Learning Forum Posts Topic Discovery and Its Application in Recommendation System

Changri Luo [1], Tingting He[2*], Xinhua Zhang [3], Zibo Zhou[4]

[1] National Engineering Research Center for e-Learning, Wuhan, China; College of Vocational and Continuing Education, Central China Normal University, Wuhan, China.
[2] Academy of Computer Science, Central China Normal University, Wuhan, China; Network Media Branch, National Language Resources Monitoring and Research Center, Wuhan, China.
[3] College of Computer ScienceWuhan Vocational College of Software and Engineering,Wuhan, China.
[4] College of Vocational and Continuing Education, Central China Normal University,Wuhan, China.

* Corresponding author. Email: tth@mail.ccnu.edu.cn.
Manuscript submitted June 13, 2014; accepted August 22, 2014.
doi: 10.17706/jsw.10.4.392-402

Abstract: What a network learner post on a network learning forum directly reflects the learners' need during from the network learning process. The network learning supporting service could be greatly improved with mining topics from the forum posts. For this purpose, this paper employs the Learner-Topic (LT) model to mine learners' posts and discover the topics. The results from the model are used to search the learners who have the same learning interest and recommend the learning resources in a recommendation system.

Key words: Network learner, topic model, LT model, post, recommendation.

1. Introduction

With the development of network technologies and applications, the network learning is growing up rapidly in real life. The web based learning resources rapidly increased, and the education information increased accordingly. Reference [1] proposes a distributed e-learning management system to use a variety of education materials scattered on the internet. Data mining is applied to study education information, as in [2],[3]. These studies are pertain to the network learners' learning habits and methods. The learning resources are recommended according to the learner's previous selections of learning resources. It only shows the learner's selections of existing learning resources, but cannot reveal what the learner really need directly. Especially, when the resources that the learners browsed are not abundant, or when learners just aimlessly or randomly clicked on these resources, the learners' selections do not reflect what they really want, and the recommendation system would give the opposite results. In addition, for new network learners, there is no information about their learning habits and the resources selection history is not available. In the above two situations, if the information is used to recommend learning resources, the cold start problem will be confronted [4], [5].

How to understand the learner's real need of the learning resources? Reference [6] extracts learners' relationship based on learning processes and learning activities to provide more authentic, personalized recommendations for group learning support. Reference [7] automatically analyzes the learning forum posts in network learning system and recommends the resources to the learners. Although processing posts content is simple, the latent semantics in posts is not fully explored. The purpose of a learner to post message(s) is to release information that reflects his/her need, such as questions and ideas, and as well hopes to get help from teachers or other learners. From the content in the posts, education providers learn what the learners really need. Nowadays, a problem we are facing is how to obtain the information accurately. It will take a lot of money and time to process posts content artificially. This paper aims at solving this problem. The purpose of this paper is to explore the topics on which the posts focus and the hiding relationship between the topics and the learners, thus to provide service for learning resource recommendation and retrieving learners with the same interest.

The outline of this paper is as follows: Section 2 reviews the related works. Section 3 describes the problem. Section 4 shows the experimental results. In Section 5 we will illustrate the application of the model to recommendation. At last, we will conclude the paper in Section 6.

2. Related Works

The network learning information mining becomes very popular in recent years. Data mining techniques, such as, decision tree, Bayesian model, classification and clustering, as in [8]-[12], are used for education information mining. All these are the results from analyzing the data produced in the learning process, such as learning habit, learning time, course work, and exam performance. Although these studies considered the interaction information among learners, learning system and teachers, they did not make full use of the learners' posts in BBS of learning systems. Posts content in BBS can directly reflect the learners' real need. For example:

1) Who has C++ language programming exercises? Please send me a copy, thanks!
2) College English is too hard for me. I can't remember words and don't know how to learn!

In the post 1), we can see the learner's idea at a glance that he wants a copy of C++ language programming exercises. A teacher understands the post with no difficulty and is easy to provide help to the learner, but how does the computer system do that? The post 2) is more difficult to understand than the post 1) and the idea of the learner is more complex than the learner in the post 1): Cannot remember the English word, do not know how to learn, it is in trouble! The computer system is very difficult to deal with the content. If the learner is a novice and no previous information about his learning is available, the personalized recommendation is difficult to provide for him [13].

In the era of information explosion, currently a major problem is how to effectively obtain the latent topic and intrinsic semantic in mass information. The post length in the learning system or the forum is usually very short, with an average word of 70, like a microblog. Reference [14] discovers the keywords in BBS by using the Influence Diffusion Model (IDM); Reference [15] adopts the Latent Dirichlet Allocation (LDA) model into microblog mining and it achieved good performance; Base on users' interested topic, [16] analyzes users' emotions by using the Network Model. The LDA model in [17] and its extension models (such as in [6],[19]) are built around the topic of data mining. The Author-Topic (AT) model in [20], one of the topic models, treats the author as latent variable to establish the link between authors and topics in document.

In this paper, in order to provide the network learning supporting services about some topics for learners, the LT (Learner-Topic) model in [21] is used to analyze the posts content sent by network learners and establish the link between the post learners and the post topics. This paper is different from [21] in using the method in recommender system.

3. Problem Description

The posts in a learning forum can be divided into the posts and the replies. They are often characterized by diverse content and varying lengths, usually relatively brief. Especially, the course forum posts have obvious characteristics as following:

1) Most of posts ask for help or questions;
2) The content of posts is relatively brief and the replies are mostly related to their theme. Moreover, only a few replies are irrelevant to the main post. Here, we define the first store of a post as the main post;
3) Sometimes, new issues related with the main post are raised in replies;
4) Expression of posts is not standardized, for example, 3ks is short for thanks in network language.

The length of a post in a learning forum is short, for example, maybe a main post is one sentence only to ask a question. And the repliers can answer the question or reference others' reply directly. Then, there is a little information available in the whole post, so we need some processing to increase semantic information. Thus, we put the main post and the replies together as a document which contains many authors. For the convenience of description, we define the following information:

Definition 1. pd is short for post-document, consisting of the text of the main post and the replies. The title of the post-document is the original title of the main post. The main poster will be treated as the first learner and the replier as one of learners in post-document; the reply text will be treated as empty and the replier will not be treated as an learner in pd if the reply is reference the main post only. Then, we will get a set of post-documents: $PD = \{pd_i\}, i = 1 \dots |PD|$.

Definition 2. Vocabulary. All distinct words from the post-document set constitute a vocabulary, $W = \{w_i\}, i = 1 \dots |W|$. For actual needs, we define noun vocabulary, $W_N = \{nw_i\}$, $i = 1 \dots |W_N|$.

Definition 3. Domain words. Post-documents, which are downloaded from network educational learning forum, are mainly about the network learning. They have the domain property of education, such as network education, expressing a specific concept, $C_D = \{c_i\}, i = 1 \dots |C_D|$.

Definition 4. Distribution description. Let us assume that a group of network learners, L, decides to post and reply a post. The model is shown in Fig. 1, a learner is chosen at random for each word in the post. Then, a topic is chosen from a distribution over the learner's interest, and the word of the post-document is generated from the chosen topic. So, the formal definition for the generative process in the LT model can be described as following:

1) For each learner x, $\theta_x \sim \text{Dirichlet}(\alpha)$;
2) For each topic k, $\Psi_k \sim \text{Dirichlet}(\beta)$;
3) For each word w_{Pdi} in the post-document pd:
   Sample a learner $x_i \sim Multinomial(L_{pdi})$;
   Sample a topic $z_i \sim Multinomial(\theta_{pdi})$;
   Sample a word $w_{pdi} \sim Multinomial\left(\Psi_{k_{pdi}}\right)$;

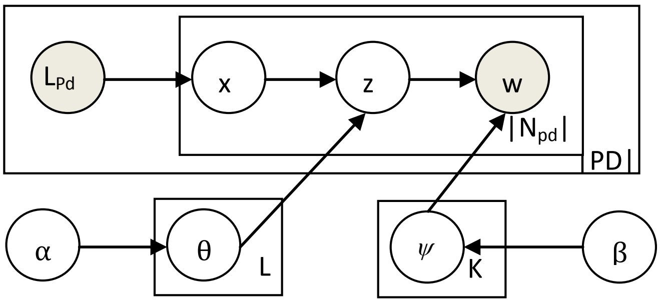

Fig. 1. The LT model.

where, pd is the post-document, $|PD|$ is the number of post-document set PD, $|N_{pd}|$ is the length of the post-document pd, L is the learner number, K is the number of topics, l_{pd} is the learners joint in the pd, x is learner, z is topic, w is word, θ is a $L \times K$ matrix indicating learner-topic distribution, Ψ is a $K \times V$ matrix indicating topic-word distribution, and α and β are the parameters of Dirichlet distribution. In this model, we use Gibbs in [22] sampling method to estimate θ and Ψ.

In this model, we have two sets of latent variables: x and z. The joint probability of the z, x assignment and the words can be factored into the following terms:

$$p(w, z, x) = p(w|z) \cdot p(z|x) = p(w|z) \cdot \big(p(z|\theta) \cdot p(x)\big) \tag{1}$$

We draw each (z_i; x_i) pair as a block, conditioned on all other variables. Since in this model, an author is chosen uniformly at random for each word that appears in the document, the sampling equation can be induced as:

$$p(z_i = k, x_i = j | w_i = m, w_{-i}, z_{-i}, x_{-i}) = \frac{C_{mj}^{WT} + \beta}{\sum_{m'} C_{mj'}^{WT} + V\beta} \cdot \frac{C_{kj}^{LT} + \alpha}{\sum_{j'} C_{kj'}^{LT} + T\alpha} \tag{2}$$

After a set of sampling processes based on the posterior distributions calculated with the above equations, we can estimate θ and Ψ using (3) and (4).

$$\theta_{kj} = \frac{C_{kj}^{LT} + \alpha}{\sum_{j'} C_{kj'}^{LT} + T\alpha} \tag{3}$$

$$\Psi_{mj} = \frac{C_{mj}^{WT} + \beta}{\sum_{m'} C_{mj'}^{WT} + V\beta} \tag{4}$$

where, θ_{kj} is the probability of topic *j* by learner *k*, Ψ_{mj} is the probability of using word m in topic *j*, C_{kj}^{LT} is the number of times that learner *k* is assigned to topic *j*, and C_{mj}^{WT} is the number of times that word m is assigned to topic *j*, excluding the current instance.

4. Experiment

4.1. Dataset

In this paper, the experiment dataset is downloaded from the learning forum of Open network education system [23], focusing on the Network Education Unified National Exam of Basic Computer Application and College English. After preprocessed dataset, the collection contains *PD*=1494 post-documents, 69943 words, and 2578 learners. Each post-document has 1.73 learners taken part in and contains 46.82 words on average.

For Topic Model, it is a hard problem to exactly define the numbers of topics. So in the experiments, we have set several different numbers of topics from 1 to 100 to evaluate the performance of our models. And In the experiment, the hyper-parameters are initialized empirically, set as: $\alpha = 0.5$, $\beta = 0.01$.

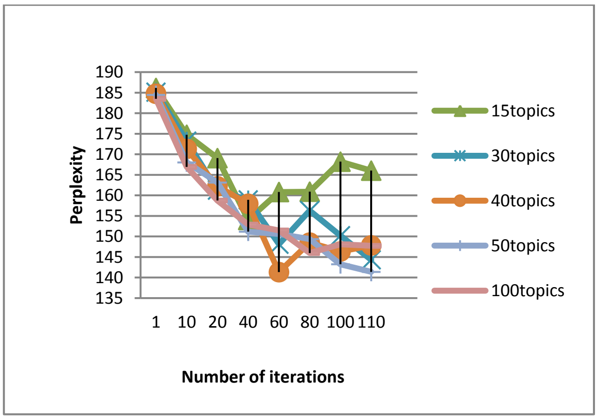

Fig. 2. The perplexity for different topic numbers.

4.2. Results of Perplexity

The perplexity is used as a standard measure to estimate the generalization performance of a probabilistic model. The value of perplexity reflects the ability of a model to generalize to unseen data. A lower perplexity score indicates better generalization performance [24], [25]. In this paper, the perplexity reflects the ability of the model to predict words with the topics.

For a test words set, (w_{pd}, l_{pd}) is a vector of post-document *pd*, $pd \epsilon PD_{test}$, and the perplexity can be defined as follows:

$$Perplexity(PD_{test}) = exp\left\{-\frac{\sum_{pd} log\,(p(w_{pd}|l_{pd}))}{\sum_{pd} N_{pd}}\right\} \tag{5}$$

$$p(w_{pd}|l_{pd}) = \sum_{k=1}^{K} p(z_k|l_{pd})p(z_k|l_{pd}) \tag{6}$$

Fig. 2 shows the results of a change of a perplexity for different topic numbers. The number of topics K has a great impact on the performance of the topic model. We try serial perplexity by the model for different numbers of topics. From Fig. 2, we can see that the perplexity decreases when the K (K=15, 30, 40, 50, 100) increases and almost fluctuates in respective convergence region after iterating 110 times. The minimum perplexity is obtained when the number of topics is set to 40, and the fluctuation is relatively stable after the convergence. So, we set the topic number K=40 in our experiments. This result coincides with the results shown in Fig. 2; the function is to converge after iterating 60 times and the fluctuation of perplexity is relatively smooth.

4.3. Examples of Topics Discovered

As shown in Fig. 3, we list the examples of three topics discovered by the LT model. Each topic is shown with the top 10 words and top 10 learners (ID), along with the probabilities. In Fig. 3, with regard topic 0, the top 10 words are "forum (luntan)", "Course work (kechengzuoye)", "education (jiaoyu)", "simulation test (moniti)", "unified examination subjects (tongkaokemu)", "fuxiti (English: exercises)", "fenzhi (English: score)", "tongkaoti (English: unified examination questions)", "difficulty (nandu)", "question (wenti)", so topic 0 may represent the semantic meaning of "Network

Topic 0		Topic 15		Topic 18	
Word	Prob	Word	Prob	Word	Prob
Forum	0.1380	Learning	0.1887	Unified examination	0.2557
Course work	0.1222	Magazine	0.1390	Be able to	0.1899
Education	0.1012	Public	0.0656	Book	0.0898
Simulation test	0.0948	Unified Examination results	0.0517	Simulation test	0.0822
Unified examination subjects	0.0548	Education	0.0426	Subject	0.0712
Exercises	0.0337	Preparing for the exam	0.0347	Syllabus	0.0606
Score	0.0306	System	0.0345	Query	0.0398
Unified examination questions	0.0295	Major	0.0334	Construe	0.0278
Difficulty	0.0295	Test	0.0325	Construe video	0.0234
Question	0.0264	Answer questions	0.0306	Exemption Conditions	0.0230
Learner	Prob	Learner	Prob	Learner	Prob
UID:926505	0.0021	UID:1511781	0.0024	UID:1122753	0.0026
UID:1561918	0.0020	UID:562710	0.0024	UID:1606967	0.0020
UID:369160	0.0020	UID:457990	0.0023	UID:1762556	0.0020
UID:1581597	0.0020	UID:1540604	0.0021	UID:347190	0.0017
UID:1174552	0.0019	UID:1273204	0.0021	UID:1543397	0.0017
UID:1214285	0.0019	UID:1720851	0.0021	UID:1435550	0.0017
UID:1274733	0.0019	UID:1512100	0.0021	UID:1210912	0.0016
UID:1211639	0.0019	UID:1763695	0.0020	UID:1923900	0.0016
UID:378658	0.0017	UID:922679	0.0020	UID:375530	0.0016
UID:1894340	0.0017	UID:1936346	0.0020	UID:1456783	0.0016

Fig. 3. Topic examples.

Education Unified National Examination forum (tongkaoluntan)" . The top 10 words of topic 15 are "xuexi (English: learning)", "magazine (zazhi)", "public (gonggong)", "unified examination results (tongkaochengji)", "education (jiaoyu)", "Preparing for the exam (beizhan)", "system (xitong)", "major (zhuanye)", "test (ceshi)", "answer questions (dayi)". It may be about the topic of "preparing for the unified examination (tongkaobeizhan)". The top 10 words of topic 18 are "unified examination (tongkao)", "be able to (hui)", "book (shu)", "simulation test (moniti)", "subject (kemu)", "syllabus (kaoshidagang)", "query (chaxun)", "construe (chuanjiang)", "construe video (chuanjiangshipin)", "exemption conditions (miankaotiaojian)".This may represent the semantic meaning of "learning or reviewing for the unified examination (tongkaofuxi)". Fig. 4 contains the post examples for example topics in Fig. 3. The mapping between topic examples in Fig. 3 and post examples in Fig. 4 has provided the connection evidence between the topics and the learners. These examples further provided evidence for that in the same topic, the learners have the same interest in learning, query or question.

Fig. 3 and Fig. 4 illustrate that the experimental performance is very good, but there are still some problems:

1) The topics about consultation, registration, learning of the unified national exam and test system maintenance are included in the corpus. However, the experiment can not accurately distinguish these topics and the results often show overlapped phenomenon. This may be related with the selected corpus in which learners' posts are often short and the words they used are overlapped, random and un-standard.
2) In the same topic, the probability is very close or even equal between the different learners and the topics. This indicates that the posts about the topic have been paid high attention to.

5. Application

5.1. Find Learners with the Same Interest

When a network learner begins to study a course, he is a new network learner, unfamiliar with the network learning, and does not know how to learn this course. So he wants to talk with the people who are learning or the people have learned before. For the network learning, the process of learning is on the network. It is difficult for a network learner to know who is learning the same course as him. How to find the learning mates for the learners? It is an essential problem to recommend learners with the same learning interest to a learner. It can be transformed to the distribution of different learners in one topic. That is, to compute the posterior probability of different learners l_k assigned to z_j while the topic z_j is known. The computing formulas are follows (7) and (8).

$$P(l_k|z_j) = \frac{P(z_j|l_k)P(l_k)}{P(z_j)} \tag{7}$$

$$P(z_j|l_k) = \theta_{kj} \tag{8}$$

This is also useful for online educators. If one topic is concerned by many learners, called common concerned hot topic, the educators will solve the common concerned issues of learners based on the common concerned hot topic, such as by recommending related learning resources to the learners. The topic examples in Fig.3 provide this application.

5.2. Recommend Learning Resource

The problem of the learning resource recommendation is that when a learner learns a course, he would like to get some learning reference materials about the course, such as reference books related to the course. This problem can be transformed into the words distribution problem in topic z condition. What is under the learner ℓ is known conditions, computing the posterior probability of word w, and ranking the probability for candidate resources. The formula is as follows:

$$p(r|\ell) = \prod_{w \epsilon r} p(w|\ell) \tag{9}$$

$$p(w|\ell) = \sum_k p(w|z_k) p(z_k|\ell) \tag{10}$$

where, r is the learning resource, ℓ is a learner. Fig. 5 is the example of learning source recommendation.

Topic 0 example:

The first post name: szs090306 UID: 1581597 TIME:2010-05-27 12:32

There seems to be a problem with the answer of online homework.

15.In the following options，（ ）is the most commonly used 3D animation making software tools.

A. 3D MAX B. Fireworks C. Photoshop D. Authorware

The answer should be A，but I select A is wrong，so I think the answer has something wrong. Similar to several questions, hope the teacher can give a reply.

The second post name: faxue-wangtt UID:926505 TIME:2010-05-27 16:10

Reply: There seems to be a problem with the answer of online homework.

Hello: excuse me, which course' online homework? Here is the unified examination forum, and the unified exam courses are not online homework.

The third post name:jichu03 UID:1196662 TIME:2010-05-27 18:47

Reply: There seems to be a problem with the answer of online homework.

Hello, student！ About the course assignments, please asking in the corresponding course forum.

Happy learning！

Topic 15 example:

The first post name:jichu04 UID: 1540604 TIME:2010-06-22 19:51

The electronic magazine second issue of "unified exam preparation technical journal" for unified examination in September, is on-line. The electronic magazine second issue of "unified exam preparation technical journal" for unified examination in September, is on-line. The electronic magazine second issue of "unified exam preparation technical journal" is online! This is the second issue for September examination. The student service center will launch 3 issues of the electronic magazine before the September examination, to help students prepare for the exam. This is a set of specialized examination review production issue, closely integrated with the unified exam coach by the Student Service Department. It includes the test analysis, unified examination hall homework, synchronous practice, classroom supplement, classic subject analysis, unified examination by-talk, the unified examination policy selection and other columns. Where, the content in "unified examination hall" is the video copy replay that the teacher answer online, which includes all the important points involved in the examination syllabus. "Synchronous practice" provides with Lectures exercises. Questions are the teachers who long-term application in the study of the exam preparation, and seize the test points. Moreover, there are professional teachers write papers to analysis examination test situation, discuss test taking skills and review methods. The electronic magazine incorporates lectures, exercises, read into an organic whole, is the mentor for students to prepare for the exam. Each issue of "unified exam preparation technical journal" include four volumes, that is "College English", "College Chinese", "higher mathematics", "computer application basis". This issue is still free to every student. Only use your own account landing the learning platform, enter the unified exam counseling column, click the publicity pictures of the electronic magazine, you can read it. Welcome the students attention!

Topic 18 example

The first post name: zy014 UID: 1762556 TIME:2009-08-29 19:55

If the unified exam simulation questions do well, can I pass the examination?

The second post name:jichuzy0014 UID: 1122753 TIME:2009-08-30 13:35

Reply: If the unified exam simulation questions do well, can I pass the examination?

Well do the simulation questions and do it right, this is to say your English is very good and I wish you success in the examination!

The third post name:ss1203 UID: 1456783 TIME:2009-08-30 20:26

Reply: If the unified exam simulation questions do well, can I pass the examination?

Hello, teacher! Whether or not the unified exam questions are the original questions of the unified exam simulation questions?

The sixth post name:jichuzy0014 UID: 1122753 TIME:2009-08-31 20:15

Reply: If the unified exam simulation questions do well, can I pass the examination?

The unified exam resources are in the forum counseling column which on the lower right corner of the forum homepage.

Fig. 4. Post examples.

Post example:
The first post name: aishishan UID:332010 TIME: 2009-04-03 10:51
Whether the English is the unified examination course?
Hello, teacher! I am a bachelor degree student of construction engineering major of Dongbei University of Finance and Economics. My question is that do I need to take part in the unified college English examination? I have checked the past each semester course planning in the same major, I have no found the college English unified examination course. There is only the computer application basis unified examination course. But, when I enrolled in learning center, the teacher told me that the college English is the unified examination course. I want to know the massage is true or not.

The second post name: xueqingteacher11 UID:1150 TIME: 2009-04-03 14:44
Reply: Whether the English is the unified examination course?
Hi, you need to take part in the college English and computer application basis unified examination. The unified examination of English and computer is one of the conditions for the graduation. However, the examination has no credit. If the unified examination courses have credit, it shows that the courses have been brought into the teaching plan, the credit is the course credit. There has nothing with the course is in teaching plan or not, please understand

Learning Source Recommendation:
Learner: UID:332010
0th: E007 1.2607030349427977E-14 《College English Tests Band Six Preparation Guide and Model Tests》
1th: C008 8.919102051068489E-22 《the Guidance on the Machine Operation and Knowledge Analysis of Computer Application Basis》
2th: E001 4.4154662848088145E-23 《College English Online》

Fig. 5. The example of learning source recommendation.

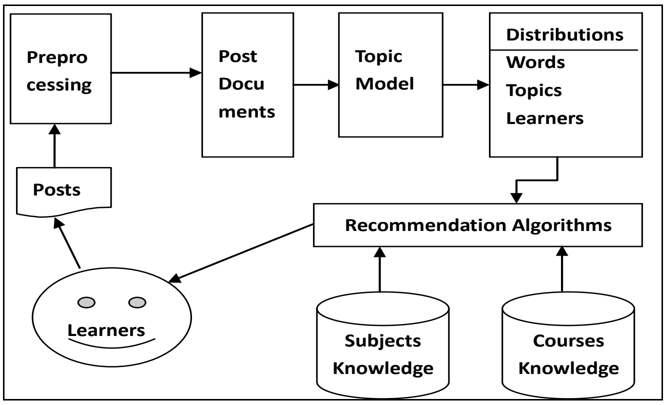

Fig. 6. Recommendation system architecture.

5.3. Recommendation System Architecture

According to the need of the online learning system users, the recommendation system workflow based on learner forum posting, is designed as Fig. 6. The whole processing of the recommendation system is as following:

Step 1. To download and preprocess the posts from the learning forum is to meet the requirement inputting file for the topic model;

Step 2. Input training data, train the topic model and estimate parameters;

Step 3. Output topic-learner and word-topic distributions based on the topic model training results;

Step 4. Appropriate recommendation algorithm is used to recommend the learners who have the same learning interest, the common concerned topic and learning resources to learners or educators.

The recommendation system needs some domain knowledge databases such as subject knowledge and course knowledge database. Here, we construct 22 subjects knowledge and 298 courses knowledge database for recommendation in our online learning platform [26]. We compute the relation between words is based on How Net [27].

6. Conclusion

In this paper, we have explored the application of data mining technology in education information processing firstly. To solve the problems existing in current studies, the LT model is applied for modeling the posts that the network learners posted on the learning forum to mine the learners focusing topic in their posts, and understand learners' need during their network learning process. So, those can help the network learners find the learning mates who have the same learning interest with themselves and implement personalized learning resources recommendation to network learners in their network learning process.

Acknowledgment

This work was supported by the major project of the National Social Science Fund (No. 12&2D223), the Major Project of State Language Commission in the Twelfth Five-year Plan Period (No.ZDI125-1), the Project in the National Science &Technology Pillar Program in the Twelfth Five-year Plan Period (No.2012BAK24B01), the Program of Introducing Talents of Discipline to Universities (No.B07042) and the NSF of Hubei Province (No.2011CDA034).

References

[1] Thongchai, K., Ryosuke, S., & Hiroshi, T. (2013). Framework for distributed e-learning management system. *Journal of Computers, 8(7)*, 1635-1647.

[2] Schiaffino, S., Garcia, P., & Amandi, A. (2008). E-teacher: Providing personalized assistance to e-learning students. *Computers and Education, 51(4)*, 1744-1754.

[3] Chang, Y. C., Kao, W. Y., & Chu C. P. (2009). A learning style classification Mechanism for e-learnig. *Computer and Education*, *53(2)*, 273-285.

[4] Schein, A. I., Popescul, A., & Ungar, L. H., & Pennock, D. M. (2002). Methods and metrics for cold-start recommendations. *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 253–260).

[5] Lam, X. N., Vu, T., Le, T. D., & Duong, A. D. (2008). Addressing cold-start problem in recommendation systems. *Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication* (pp. 208–211).

[6] Xin, W., Qimanguli, J., & Toshio, O. (2011). Analyzing learners' relationship to improve the quality of recommender system for group learning support. *Journal of Computers*, 254-262.

[7] Daniil, C., Dascalu, M., & Stefan, T. M. (2012). Automatic forum analysis a thorough method of assessing the importance of posts, discussion threads and of users' involvement. *Proceedings of the International Conference on Web Intelligence, Mining and Semantics*.

[8] Hurley, T., & Weibelzahl, S. (2007). Using MotSaRT to support on-line teachers in student motivation. *Proceedings of the European Conference on Technology-Enhanced Learning* (pp. 101-111).

[9] Vranic, M., Pintar, & D., Skocir, Z., (2007). The use of data mining in education environment. *Proceedings of the International Conference on Telecommunications* (pp. 243-250).

[10] Hübscher, R., Puntambekar, S., & Nye, A. (2007). Domain specific interactive data mining. *Proceedings of the Workshop on Data Mining for User Modeling, at the 11th International Conference on User Modeling*(pp. 81-90).

[11] Ranjan, J., & Khalil, S. (2008). Conceptual framework of data mining process in management education in India: An institutional perspective. *In Information Technology Journal*, 16-23.

[12] Chen, C., & Chen, M., (2009). Mobile formative assessment tool based on data mining techniques for supporting web-based learning. *Computer and Education Journal*, *52(1)*, 256-273.

[13] Zhang, Z. K., Liu, C., Zhang, Y. C., & Zhou, T. (2010). Solving the cold-start problem in recommender systems with social tags.

[14] Tsuda, K., & Thawonmas, R. (2005). Keyword discovery by measuring influence rates on bulletin board services. *International Cost Engineering Council*, 2148-154.

[15] Li, D. F. (2012). Mining topic-level opinion influence in microblog. *Proceedings of the Information and Telecommunication Technology Center (ITTC) Advances Knowledge and Creates Innovative Technologies in Telecommunications, Information systems, Bioinformatics, and Radar* (pp. 1562-1566).

[16] Yu, X., Wei, X., & Lin, X. (2010). Algorithms of BBS opinion leader mining based on sentiment analysis. *Waldorf Institute of Southeastern Michigan*, 360–369.

[17] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 993-1022.

[18] Tang, J., Zhang, J., Jin, & R. M. (2011). Topic level expertise search over heterogeneous networks. *Mach Learn*, 211-237.

[19] Wang, L. D., Wei, B. G., & Yuan, J. (2011). Topic discovery based on LDA_col model and topic significance re-ranking. *Journal of Computers*, *6(8)*, 1639-1647.

[20] Rosen, Z. M., Griths, T., Steyvers, M., & Smyth, P. (2004). The author-topic model for authors and documents. *Unemployment Insurance Agency*, 487-494.

[21] Luo, C., He, T., & Zhang, X. (2014). Network learning forum posts topic discovery. *Proceedings of the Transactions on Information and Communication Technologies* (pp. 363-370).

[22] Griffths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences* (pp. 5228-5235).

[23] Teaching platform. Retrieved, from http://learn.open.com.cn/Login.aspx.

[24] Chen, X., Hu, X. H., Zhou, Z. N., & Lu, C. M., Rosen, G., He, T. T., & Park, E. K. (2010). A probabilistic topic-connection model for automatic image annotation. *Proceedings of the Conference on Information and Knowledge Management* (pp. 809-908).

[25] He, T. T., & Li, F. (2012). Semantic knowledge acquisition from blogs with tag-topic model. *China Communications*, *9(3)*, 38-48.

[26] Learning platform. Retrieved, from http://www.hznu.cn/.

[27] Liu, Q., & Li, S. J. (2002). *Word Similarity Computing Based on How-net, Computational Linguistics and Chinese Language Processing*.

Changri Luo is currently a Ph.D. candidate and a teacher at Central China Normal University. His research interests include natural language processing and artificial intelligence.

Tingting He received her Ph.D. degree in literature from Central China Normal University in 2003. She is currently a professor at the Department of Computer Science at Central China Normal University. Her research interests include natural language processing and artificial intelligence.

Xinhua Zhang is currently a teacher at the Computer Science College of Wuhan Vocational College of Software and Engineering. Her research interests include natural language processing and artificial intelligence.

Zibo Zhou is currently a teacher at the College of Vocational and Continuing Education, Central China Normal University.