Content Filtering Based on Keyword Map

Xin Wan[1, a], Neil Rubens[2],Toshio Okamoto[2] and Yan Feng[1]

[1]School of Management, Wuhan Textile University, Wuhan, China

[2]Graduate School of Information System, The University of Electro-Communications, Tokyo, Japan

[a]xwan-wtu@sohu.com

Keywords: user profiling; keyword map; recommender system; group learning

Abstract. This paper presents the content filtering based on keyword map. Keyword maps represent knowledge of learner by capturing relations between terms that the learner has been exposed to. Keyword maps allow increasing both the relevance and complement of learning resources recommendation. The experimental results show that the proposed method is better suited for e-learning settings our method, and achieves a higher accuracy than common recommendation methods such as collaborative and keyword based approaches.

Introduction

Nowadays personalization is regarded as crucial in e-learning settings. Many e-learning systems have applied recommender systems to support learning resources personalization, such as the paper recommendation proposed by Tang et al. [1], the learning activities recommendation developed by Drachsler et al. [2], the learning process recommendation proposed by Wan et al. [3,4,5].

Content-based recommender systems always use keyword vectors to carry out recommendations. The process of content-based recommender systems is as follows: (1) identifying the item contents, (2) identifying the user's interests and (3) using the methods used to match them. Therefore, in order to design the accuracy content-based recommender system, it is necessary to consider the following problems, (1) how to represent a content of items (content profiles), (2) how to create user profiles that stand for users' preferences.

In this paper, we propose keyword map based recommendation method. Keyword maps represent knowledge of learner by capturing relations between terms that the learner has been exposed to. Keyword maps are constructed based on the learner's user profile, which consists of visited learning resources, and learning processes in a web based learning system.

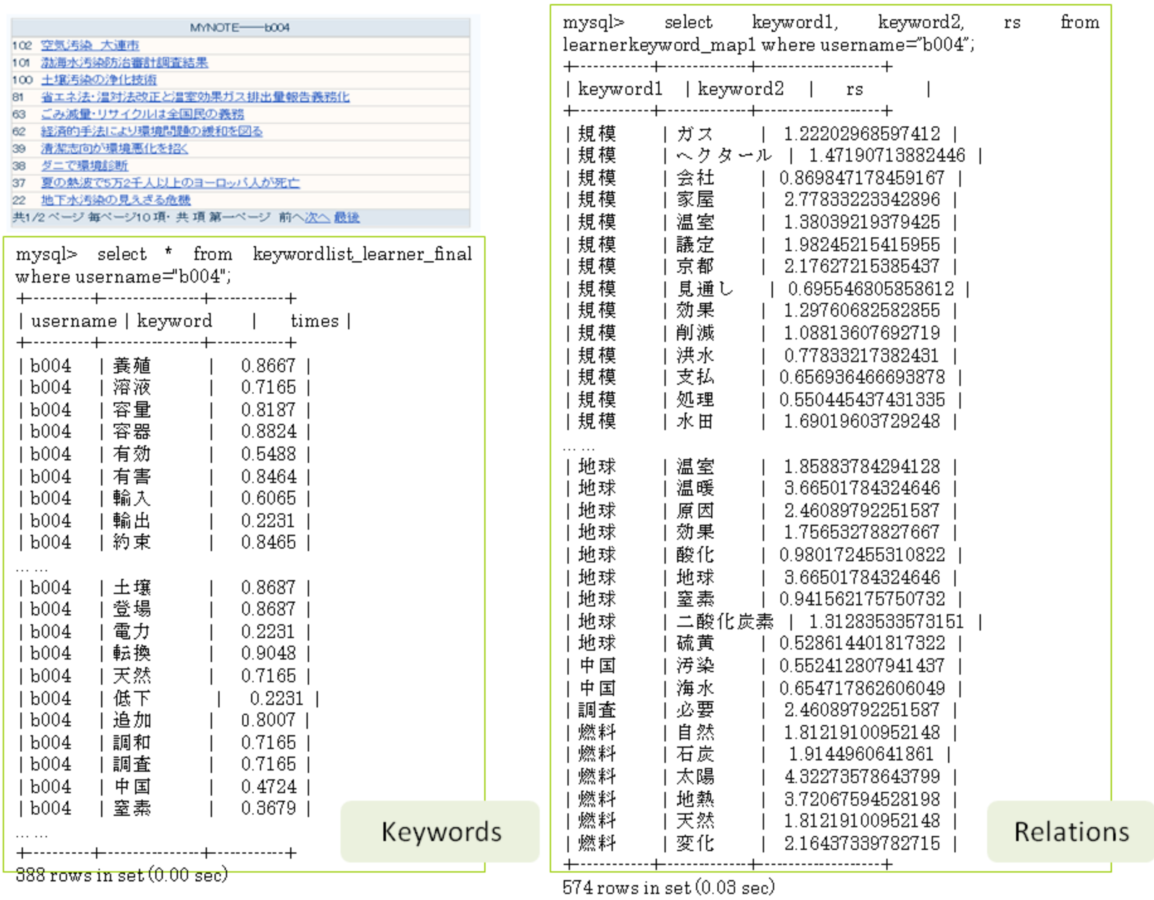
Approach

In e-learning settings, keyword vector-based methods are not necessarily effective, since recommendation of learning resources depends not only on the keywords, but also on existing knowledge of learner and its relations to a resource considered for recommendation (e.g. appropriate difficulty, subject area, course plan, etc.). To overcome these limitations we propose keyword map based recommendation method [1]. The approach utilize keyword map to provide a mechanism for representing the information of text. Each keyword in a keyword map is used to represent a subject or object of a sentence. Keyword maps are constructed based on the learner's user profile, which consists of visited learning resources, and learning processes in a web based learning system. Content filtering based on keyword map (Fig.1, Fig.2, Fig.3) consists of two parts: (1) generating learner's keyword map based profile and (2) Jaccard recommendation.

Using keyword map on the construction of user profiles: the system creates learners' keyword map based profiles automatically based on their learning processes implicitly (Tab.1).

Jaccard recommendation: we use a fuzzy version of the standard Jaccard similarity measure between the keyword map of learner profile and keyword map of each learning resources to decide the recommendation.

The structure of keyword map allows restricting the search space and inferring missing information. The intuition behind content filtering based on keyword map is that the information represent by keyword map can be used to estimate missing preferences.



MYNOTE——b004

102 空気汚染 大連市
101 渤海水汚染防治審計調査結果
100 土壌汚染の浄化技術
81 省エネ法･温対法改正と温室効果ガス排出量報告義務化
63 ごみ減量･リサイクルは全国民の義務
62 経済的手法により環境問題の緩和を図る
39 清潔志向が環境悪化を招く
38 ダニで環境診断
37 夏の熱波で5万2千人以上のヨーロッパ人が死亡
22 地下水汚染の見えざる危機
共1/2 ページ 毎ページ10 項･ 共 項 第一ページ 前へ次へ 最後

```
mysql> select * from keywordlist_learner_final where username="b004";
+---------+--------------+-----------+
| username | keyword     |    times |
+---------+--------------+-----------+
| b004   | 養殖         |   0.8667 |
| b004   | 溶液         |   0.7165 |
| b004   | 容量         |   0.8187 |
| b004   | 容器         |   0.8824 |
| b004   | 有効         |   0.5488 |
| b004   | 有害         |   0.8464 |
| b004   | 輸入         |   0.6065 |
| b004   | 輸出         |   0.2231 |
| b004   | 約束         |   0.8465 |
… …
| b004   | 土壌         |   0.8687 |
| b004   | 登場         |   0.8687 |
| b004   | 電力         |   0.2231 |
| b004   | 転換         |   0.9048 |
| b004   | 天然         |   0.7165 |
| b004   | 低下         |    0.2231 |
| b004   | 追加         |   0.8007 |
| b004   | 調和         |   0.7165 |
| b004   | 調査         |   0.7165 |
| b004   | 中国         |   0.4724 |
| b004   | 窒素         |   0.3679 |
… …
+---------+--------------+-----------+
388 rows in set (0.00 sec)
```

```
mysql> select keyword1, keyword2, rs from learnerkeyword_map1 where username="b004";
+----------+-----------+-----------------+
| keyword1  | keyword2  |    rs          |
+----------+-----------+-----------------+
| 規模      | ガス       |  1.22202968597412 |
| 規模      | ヘクタール |  1.47190713882446 |
| 規模      | 会社       | 0.869847178459167 |
| 規模      | 家屋       |  2.77833223342896 |
| 規模      | 温室       |  1.38039219379425 |
| 規模      | 議定       |  1.98245215415955 |
| 規模      | 京都       |  2.17627215385437 |
| 規模      | 見通し     | 0.695546805858612 |
| 規模      | 効果       |  1.29760682582855 |
| 規模      | 削減       |  1.08813607692719 |
| 規模      | 洪水       |  0.77833217382431 |
| 規模      | 支払       | 0.656936466693878 |
| 規模      | 処理       | 0.550445437431335 |
| 規模      | 水田       |  1.69019603729248 |
… …
| 地球      | 温室       |  1.85883784294128 |
| 地球      | 温暖       |  3.66501784324646 |
| 地球      | 原因       |  2.46089792251587 |
| 地球      | 効果       |  1.75653278827667 |
| 地球      | 酸化       | 0.980172455310822 |
| 地球      | 地球       |  3.66501784324646 |
| 地球      | 窒素       | 0.941562175750732 |
| 地球      | 二酸化炭素 |  1.31283533573151 |
| 地球      | 硫黄       | 0.528614401817322 |
| 中国      | 汚染       | 0.552412807941437 |
| 中国      | 海水       | 0.654717862606049 |
| 調査      | 必要       |  2.46089792251587 |
| 燃料      | 自然       |  1.81219100952148 |
| 燃料      | 石炭       |   1.9144960641861 |
| 燃料      | 太陽       |  4.32273578643799 |
| 燃料      | 地熱       |  3.72067594528198 |
| 燃料      | 天然       |  1.81219100952148 |
| 燃料      | 変化       |  2.16437339782715 |
+----------+-----------+-----------------+
574 rows in set (0.03 sec)
```


Fig.1 Example of keyword map based learner profile (in database)

This is User b004's keyword map based learner profile.

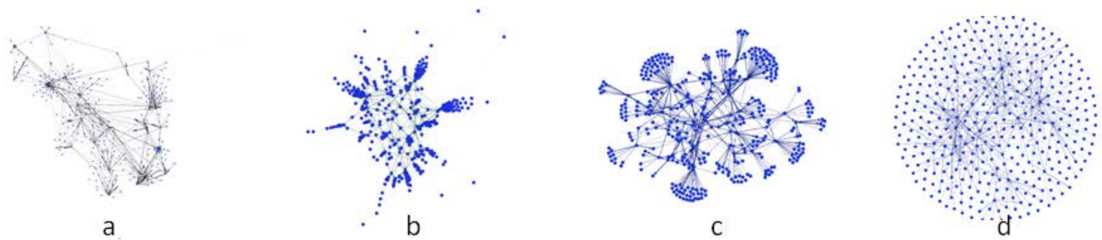


Fig.2 Example of keyword map based learner profile (Visualization)

This is the real keyword map profile of learner "b004" (keyword: 388; relation: 574). The visualization is based on Gephi (www.gephi.org). The layouts of keyword map are (a) OpenOrd; (b) Force Altlas; (c) Yifan Hu; (d) Fruchterman Reingold.

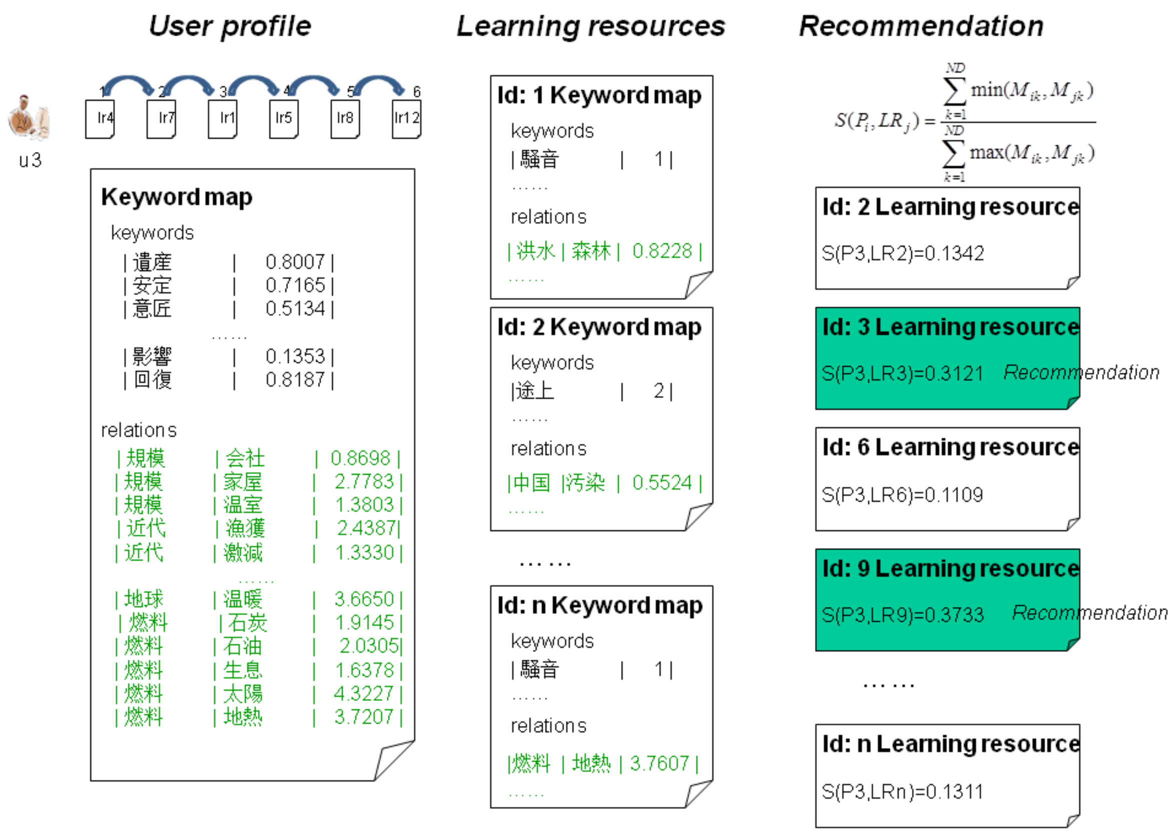


Fig.3 Example of content filtering based on keyword map

Tab.1 Learner Profile Generation Algorithm

Algorithm: Keyword Map based Learner Profile Generation
Input: a Learning Process $LP_{ui} = \{lp_{ui}^{j}\}_{j=1}^{n}$ and a set of keyword maps of learning resources
$Learner_u \leftarrow$ Empty map { is a map storing keywords and relations of them used to construct the learner profile}
for j = 1 to n do
  get $Keyword\_Map = <k_a, r_b>$ of lp_{ui}^{j}
  let $\alpha = \exp(-j/n)$
    if k_a is not found in $Learner_u$ then
    Add $\alpha \times k_a$ to $Learner_u$
    else do
    Change k_a for $\max(k_a \in Learner_u, \alpha \times (k_a \in lp_{ui}^{j}))$ in $Learner_u$
    end if
    if r_b is not found in $Learner_u$ then
    Add $\alpha \times r_b$ to $Learner_u$
    else do
    Change r_b for $\max(r_b \in Learner_u, \alpha \times (r_b \in lp_{ui}^{j}))$ in $Learner_u$
    end if
end for
Output: $Learner_u$

Note: k is keyword; r is relation.

Experiment Evaluation

Aim

The performance of the proposed recommendation approach is examined and we address the following questions:

(1) How does our approach compare with conventional content-based filtering (keyword based)?

(2) Does the proposed keyword map based learner profile generation algorithm give the better quality to recommendation?

Participant and procedure

To evaluate the performance of the proposed recommendation approaches, two approaches (keyword map content based filtering and the conventional content based filtering) were implemented in a web-based learning system [4, 5]. And 10 learners that were students of masters' degree and postgraduate studied as a group participated in the experiment. A total of 177 learning resources about "the problem of environment" were provided. The test was continued one month. And the learners evaluated the first ten of ranking lists of the recommendations.

Tab.2 Comparison of Different Methods

Recommendation technology	Main technology	Algorithm	Acquisition of user profile	R
CBF	Keyword based	Jaccard	explicit (1-to-5 scale)	69.11
KM-CBF	Keyword map based	Jaccard	Implicit (learning process)	78.06

Note: The bigger the value, the better the performance.

Tab.3 Comparison of Different Methods

Acquisition of learner profiles		R
input	output	
(a) learning resources' keyword map learning process (consider learning step)	Keyword map based learner profiles	78.06
(b) learning resources' keyword map learning process (not consider learning step)	Keyword map based learner profiles	71.34

Note: Recommendation technology is Content filtering based on keyword map (basic algorithm:Jaccard). Approach (a)'s description is in Tab.1. Learning step is not considered in approach (b), that is to say, in Tab.1 learner profile generation algorithm, α≡1.

Results

Accuracy: We used Rank Accuracy Metrics proposed by Breese et al..[6]. And this method evaluates an unbounded recommendation list that potentially contains all the items in the catalog.

From Tab.2 (where MD=0.68(a) and MD=0.68(b)), we can see that content filtering based on keyword map outperforms the conventional content-based filtering. In addition, as Tab.3 (where MD=0.56(a) and MD=0.68(b)) shows, learning step is the important element in learning process.

Coverage: We compare the CBF and KM-CBF in terms of completion coverage vs. the size K of the recommendation list. The results are presented in Fig.4 where MD=0.0128(KM-CBF, k=5), MD=0.0164(KM-CBF, k=10), MD=0.0145(CBF, k=5), MD=0.0232(CBF, k=10). KM-CBF outperforms the CBF in all cases. The reason is the structure of KM-CBF allows restricting the search space and inferring the missing information.

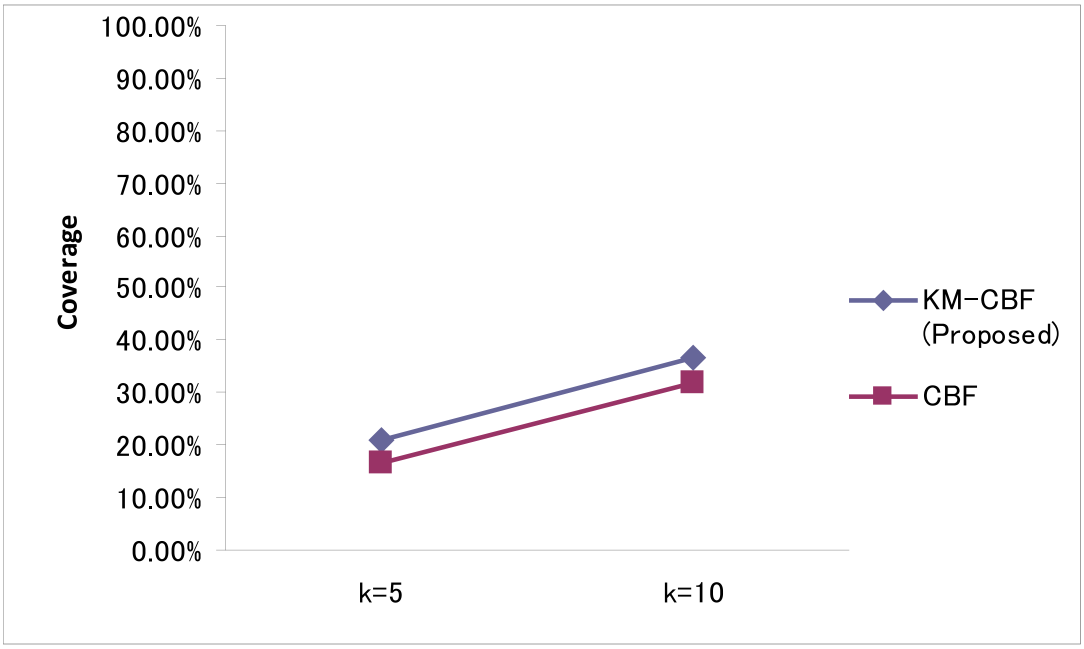


Fig.4 The result of coverage

The result implies that proposed approach (KM-CBF: content filtering based on keyword map) allows restricting the search space and inferring missing information. The information represented by keyword map can be used to estimate missing preferences. Higher coverage is better.

Discussion and Conclusions

From the educational aspect, the mechanism of content filtering based on keyword map (KM-CBF) proposed in this paper as inquiry reasoning is used to guide the learner to more interesting or useful activities during learning. Inquiry is the key part of constructivist learning. System explains ratings in terms of informative features and explains features in terms of examples based on properties. In the domain of learning, the features such as learners' characteristics and types of instructional materials have been identified on account of the success of constructing knowledge. It mainly aids in inquiry reasoning based on transfer of learning. It assists to create stories and analogy making by using individuals' learning processes. In this approaches, we use the module "learner's keyword map based profile generator" to create stories based on transfer of learning to express individuals' experiences and backgrounds. And analogy making is based on the module "Jaccard recommender module" to encourage the learner to arrive at his or her version of the truth.

From technical aspect, (1) Keyword map describes content of each learning resource and knowledge of each learner existing using keywords and various relations among them. The intuition behind content filtering based on keyword map is that the information represent by keyword map can be used to estimate missing preferences. And this structure allows restricting the search space and inferring missing information. Thus keyword maps help to increase both the completion and relevance of learning resources recommendation. (2) The learner profile generation based on learning process is effective approach, and especially, learning step is an important element in learning process.

This study proposed a new learner profiling to improve and upgrade the effectiveness of content-based learning resources recommendation.

Acknowledgements

This research has been supported by Hubei Province Fund of China (B2013207, 13g296).

References

[1] T. Tang and G. McCalla: Smart Recommendation for an Evolving E-Learning System: Architecture and Experiment, International Journal on E-Learning, 4(1), p.105-129, 2005.

[2] H. Drachsler, H.G.K. Hummel, and R. Koper: Personal recommender system for learners in lifelong learning networks: requirements, techniques and model, International Journal of Learning Technology, 3(4), p.404-423, 2008.

[3] X. Wan, Q. Jamaliding, F. Anma, T. Okamoto: Applying Keyword Map Based Learner Profile to a Recommender System for Group Learning Support, Proc. International Workshop on Education Technology and Computer Science, Vol.1, p.3-6, 2010.

[4] X. Wan, T. Okamoto: Utilizing learning process to improve recommender system for group learning support, Neural Computing and Applications, 20(5), p.611-621, 2011.

[5] X. Wan, Q. Jamaliding, T. Okamoto: Analyzing learners' relationship to improve the quality of recommender system for group learning support, Journal of Computers, 6(2), p.254-262, 2011.

[6] J. S. Breese, D. Heckerman, and C. M. Kadie: Empirical analysis of predictive algorithms for collaborative filtering, Proc. UAI, pp. 43- 52, 1998.