A Semantic Recommender System for Learning Based on Encyclopedia of Digital Publication

Mao Ye[1,2,3], Lifeng Jin[2], Zhi Tang[1,2], and Jianbo Xu[2]

[1] Peking University, Beijing, China
xjtuyemao@163.com
[2] State Key Laboratory of Digital Publishing Technology
(Peking University Founder Group Co. LTD.), Beijing, China
[3] Postdoctoral Workstation of the Zhongguancun Haidian Science Park, Beijing, China

Abstract. Digital publication is a useful and authoritative resource for knowledge and learning. How to use the knowledge in digital publication resources so as to enhance learning is an interesting and important task. Most of the recommender systems use users' preferences or history data for computation, which cannot solve the problems such as cold start, scarcity of history data or preferences data. A semantic recommender system is presented in this paper based on encyclopedic knowledge from digital publication resources, without considering history data or preferences data for learning the knowledge of a specific domain. Semantic relatedness is computed between concepts from the encyclopedia. The related concepts are recommended to users when one concept is reviewed. The method shows potential usability for domain-specific knowledge service.

Keywords: recommender system, digital publication, semantic relatedness.

1 Introduction

Digital publication has become one of the primary means for information circulation. It includes the digital publication of e-books, EPUBs, digital newspaper, digital magazines, digital encyclopedia, digital yearbook, and so on. The information in digital publication resource is normally useful and authoritative. It is an interesting and important task to associate different types of knowledge in digital publication resources in order to enhance learning. Since information retrieval (in terms of searching for relevant learning resources) is a pivotal activity in TEL (Technology Enhanced Learning), the deployment of recommender systems has attracted increased interest [1][2]. However, most of the recommender systems use users' preferences or history data for computation [3][4], which cannot solve the problems such as cold start, scarcity of history data or preferences data. Encyclopedia as digital publication is a kind of reference work containing a summary of information from either all domains of knowledge or a particular domain of knowledge in the format of articles or entries. It focuses on factual information to cover the concept for which the label stands. One purpose of an encyclopedia is to collect items around the globe and to organize the

C. Stephanidis (Ed.): HCII 2014 Posters, Part II, CCIS 435, pp. 189–194, 2014.

most important concepts into domains. A semantic recommender system is presented in this paper based on encyclopedia, without considering history data or preferences data for learning the knowledge of a specific domain.

2 Problem Domain

A typical encyclopedia mainly consists of concepts which usually describe and explain entities or knowledge points in its predestined domain. The knowledge contained in an encyclopedia may be denoted as $K = \{A, B\}$, where A is the set of labels and explanations in plain text of all concepts and B is other information than A in the encyclopedia, such as figures and pictures. The knowledge A may be represented as $O = \{o_1, o_2, \cdots, o_n\}$, where $o_i, i = 1, \cdots, n$ represents a concept. Each concept contains a label, which would be the name of the concept, and an explanation, which would be a short piece of text which describes the concept. Other concepts may be mentioned in the explanation of one concept. We suppose that the label of concepts in the encyclopedia is unique and encyclopedia contains the important concepts in one domain. Let $X = \{x_i, i = 1, \cdots, n\}$ be the label set in the encyclopedia where x_i is the label of the concept o_i . Let $Y = \{y_i, i = 1, \cdots, n\}$ be the explanation set where y_i is the explanation of the concept whose label is x_i . Therefore, it is known that $A = \{(x_i, y_i), i = 1, \cdots, n\}$ where x_i and y_i is the label and the explanation of the concept o_i respectively. When a concept o_i is reviewed by user, it is necessary to compute and recommend the related concept set $O_i^{'} = \{o_j \mid o_j \in O\}$ so that the user can learn or understand the concept o_i more effectively.

3 Semantic Recommender System

3.1 Semantic Relatedness Computing

Semantic relatedness is computed in this paper with concepts' labels $X = \{x_i, i = 1, \cdots, n\}$ and explanations $Y = \{y_i, i = 1, \cdots, n\}$ extracted from encyclopedia. Since one concept may be mentioned in other concepts' explanations and such concepts are often related with the original concept, we can extract the semantic relationship between the concepts by finding the relation among the text in explanations. The process to compute the relatedness can be summarized in the following steps.

Step 1: Extract concepts $O = \{o_1, o_2, \cdots, o_n\}$ from encyclopedia.

Step 2: For all concepts of O, extract concepts' labels $X = \{x_i, i = 1, \cdots, n\}$ and concepts' explanations $Y = \{y_i, i = 1, \cdots, n\}$.

Step 3: Compute the explicit relation from one concept o_i to another concept o_j by the equation $f_E(i, j) = \frac{\alpha(f_P(i, j) + f_N(i, j))}{1 + \alpha}$, where $\alpha \geq 1$ is a factor to control the relation strength. In the equation above, $f_P(i, j) = \frac{2}{1 + \exp(-\beta\mu)} - 1$ and $f_N(i, j) = \frac{f_P(j, i)}{\alpha}$, where $\beta > 0$ is a control factor and $\mu \geq 0$ is the number of occurrence of x_j in y_i.

Step 4: Build an explicit relation graph $G = (V, E)$ with the concepts O as vertexes and explicit relations between the concepts as edges. The edge's weight c_{ij} from the vertex o_i to o_j is set to $f_E(i, j)$.

Step 5: Build an auxiliary graph $G' = (V', E')$ according to explicit relation graph $G = (V, E)$, where $V = V'$. The weight of the edge from the vertex o_i to o_j in the graph G' is set to $c'_{ij} = -\ln c_{ij}$, where c_{ij} is the weight of the edge from the vertex o_i to o_j in the graph G.

Step 6: Compute the implicit relation from one concept o_i to another concept o_j by the equation $f_I(i, j) = \exp(-c(p^*_{ij}))$, where $c(p^*_{ij})$ is the cost of the shortest path p^*_{ij} from the vertex o_i to o_j in the graph G'. A classic algorithm to compute the shortest path is Dijkstra algorithm [5].

Step 7: Set the semantic relatedness from the concept o_i to the concept o_j with $f_E(i, j)$ if $f_E(i, j) \geq f_I(i, j)$; otherwise, set the semantic relatedness with $f_I(i, j)$.

3.2 Recommender System Based on Semantic Relatedness

The main process of the recommender system based on semantic relatedness is as Figure 1. Firstly, the concepts $O = \{o_1, o_2, \cdots, o_n\}$ of a specific domain are extracted from the encyclopedia selected for the domain. The concepts consist of labels and explanations. Ontology of the domain is then created with the concepts and the information extracted from other resources. Before the ontology can be built for a specific domain, domain experts are invited to create a schema for the domain firstly. Then the values of properties for the concepts as instances are extracted from the digital publication resources or the web resources for the ontology. The detailed

process to create the ontology is well beyond the scope of the paper. Some methods or techniques can be found from the references [6][7][8]. After the ontology is built, the important concepts are saved as instances and the information or values are associated with the concepts. The semantic relatedness is then computed for the concepts $O = \{o_1, o_2, \cdots, o_n\}$ and will be saved as a matrix M in the system, where M_{ij} represent the semantic relatedness from o_i to o_j. When a concept o_i is reviewed by users, the row of o_i in M are extracted and the concepts o_j is sorted by the value of M_{ij} in descending order. Then $O_i^{'} = \{o_j \mid o_j \in O\}$ is generated by selecting the top N concepts and are recommended to the users. The concepts in $O_i^{'}$ have better semantic relation with the concept o_i so that the user can learn or understand the concept o_i more effectively.

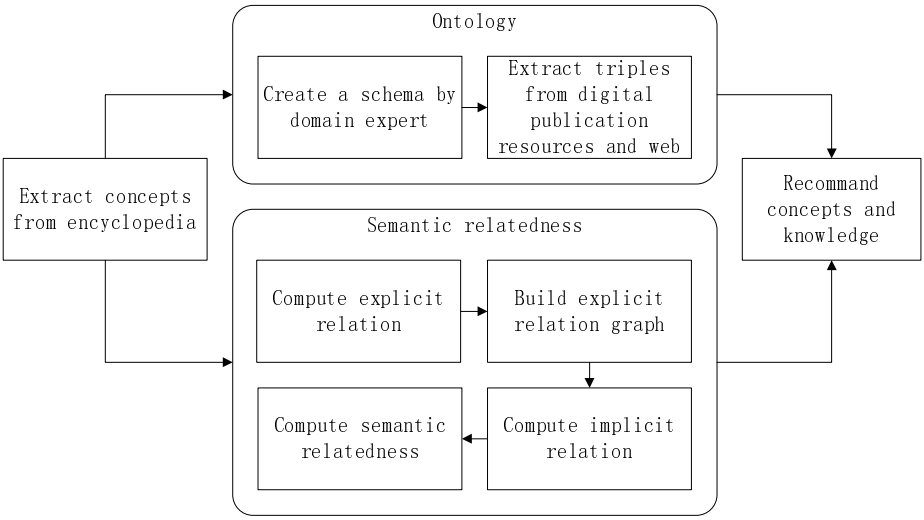

Fig. 1. The main process of the recommender system

4 A Case Study

A case of the concept Qinshihuang (秦始皇) is shown in this section to describe the early study results of the method. The concepts are extracted from three encyclopedias in the domain of history with the titles of the books "Encyclopedia of China, Chinese History I", "Encyclopedia of China, Chinese History II", and "Encyclopedia of China, Chinese History III" [9]. The number of concepts selected is 2392 for the

experiment. Ontology is then built for the concepts selected. Semantic relatedness is computed between all the concepts by the method proposed with the parameters $\alpha = 2$ and $\beta = 1$. Because the semantic relatedness is computed by encyclopedias and there are normally encyclopedias of digital publish resources written in different languages, the method can be used in different linguistic environments. This is not the case for other semantic relatedness algorithms such as WordNet-based method [10], ESA[11] and WikiRelate![12]. The value of semantic relatedness is saved in the matrix M. Let o_i be the concept Qinshihuang, we get the row of o_i in M which represents the semantic relatedness from o_i to the other concepts. Top 50 concepts are selected within which 15 are shown in the table 1. The value in the field "Relation" is obtained from the matrix M. It is shown from the table that concepts related with Qinshihuang are found effectively. These concepts will be recommended to users when they review the concept Qinshihuang so that they can learn or understand it more effectively.

Table 1. Concepts recommended for the concept Qinshihuang

Concept	Relation	Concept	Relation	Concept	Relation
秦朝	0.936765502	吕不韦	0.661768490	赵高	0.462117157
李斯	0.932303602	陈胜、吴广起义	0.639763023	云梦秦律	0.462117157
蒙恬	0.836600870	黔首	0.629420632	匈奴	0.462117157
战国	0.796724106	秦郡	0.609794189	灵渠	0.462117157
秦国	0.757471222	秦二世胡亥	0.609794189	奴隶	0.398362041
…	…	…	…	…	…

5 Conclusions

Digital publication whose resources are useful and authoritative has become one of the primary means for information circulation. A semantic recommender system is proposed in this paper based on encyclopedia of digital publication for e-learning. Semantic relatedness is computed between concepts extracted from the encyclopedia for a domain. The concepts will be recommended to users for the original concept according to the semantic relatedness between them so that users can learn or understand the original concept more effectively. The recommender system works without considering users' history data or preferences data. This is the first stage investigation. The case shows its potential usability for domain-specific knowledge service. We will combine the knowledge in the constructed ontology to improve the effectiveness of the method in the future work.

Acknowledgment. The work was funded by China Postdoctoral Science Foundation and Beijing Postdoctoral Science Foundation of China.

References

1. Manouselis, N., Drachsler, H., Verbert, K., Duval, E.: Recommender Systems for Learning, pp. 1–20. Springer, Heidelberg (2013)
2. Verbert, K., Manouselis, N., Ochoa, X., Wolpers, M., Drachsler, H., Bosnic, I., Duval, E.: Context-aware Recommender Systems for Learning: A Survey and Future Challenges. IEEE Transactions on Learning Technologies 5(4), 318–335 (2012)
3. Bobadilla, J., Ortega, F., Hernando, A., Gutierrez, A.: Recommender Systems Survey. Knowledge-Based Systems 46, 109–132 (2013)
4. Carrer-Neto, W., Hernandez-Alcaraz, M.L., Valencia-Garcia, R., Garcia-Sanchez, F.: Social Knowledge-based Recommender System, Application to the Movies Domain. Expert Systems with Applications 39(12), 10990–11000 (2012)
5. Dijkstra, E.W.: A Note on Two Problems in Connexion with Graphs. Numerische Mathematik 1, 269–271 (1959)
6. Zouaq, A., Nkambou, R.: Evaluating the Generation of Domain Ontologies in the Knowledge Puzzle Project. IEEE Transactions on Knowledge and Data Engineering 21(11), 1559–1572 (2009)
7. Ameen, A., Khan, K.U.R., Rani, B.P.: Creation of Ontology in Education Domain. In: 4th International Conference on Technology for Education, pp. 237–238. IEEE Press, New York (2012)
8. Gaeta, M., Orciuoli, F., Paolozzi, S., Salerno, S.: Ontology Extraction for Knowledge Reuse: the e-Learning Perspective. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 41(4), 798–809 (2011)
9. Encyclopedia of China, Chinese History I, II, III, `http://shuyuan.apabi.com`
10. Budanitsky, A., Hirst, G.: Semantic Distance in WordNet: An Experimental, Application-oriented Evaluation of Five Measures. In: Workshop on WordNet and Other Lexical Resources, Second Meeting of the North American Chapter of the Association for Computational Linguistics, Pittsburgh, USA (2001)
11. Gabrilovich, E., Markovitch, S.: Computing Semantic Relatedness Using Wikipedia-based Explicit Semantic Analysis. In: 20th International Joint Conference on Artificial Intelligence, pp. 1606–1611. Morgan Kaufmann Publishers Inc., San Francisco (2007)
12. Strube, M., Ponzetto, S.P.: WikiRelate! Computing Semantic Relatedness Using Wikipedia. In: 21st National Conference on Artificial Intelligence, vol. 2, pp. 1419–1424. AAAI Press (2006)