

Article

Recommender System for E-Learning Based on Semantic Relatedness of Concepts

Mao Ye ^{1,2,*}, Zhi Tang ¹, Jianbo Xu ¹ and Lifeng Jin ¹

¹ State Key Laboratory of Digital Publishing Technology (Peking University Founder Group Co. Ltd.), Beijing 100089, China; E-Mails: tommie@founder.com.cn (Z.T.); xu.jb@founder.com.cn (J.X.); jinlf@founder.com.cn (L.J.)

² Postdoctoral Workstation of the Zhongguancun Haidian Science Park, Beijing 100089, China

* Author to whom correspondence should be addressed; E-Mail: yemao.apb@founder.com.cn; Tel.: +86-10-8217-9657; Fax: +86-10-8217-9500.

Academic Editors: Lawrence J. Henschen and Ning Li

Received: 28 December 2014 / Accepted: 27 July 2015 / Published: 4 August 2015

Abstract: Digital publishing resources contain a lot of useful and authoritative knowledge. It may be necessary to reorganize the resources by concepts and recommend the related concepts for e-learning. A recommender system is presented in this paper based on the semantic relatedness of concepts computed by texts from digital publishing resources. Firstly, concepts are extracted from encyclopedias. Information in digital publishing resources is then reorganized by concepts. Secondly, concept vectors are generated by skip-gram model and semantic relatedness between concepts is measured according to the concept vectors. As a result, the related concepts and associated information can be recommended to users by the semantic relatedness for learning or reading. History data or users' preferences data are not needed for recommendation in a specific domain. The technique may not be language-specific. The method shows potential usability for e-learning in a specific domain.

Keywords: recommender system; digital publishing; semantic relatedness

1. Introduction

Digital publishing resources include e-books, digital newspapers, digital magazines, digital encyclopedias, digital yearbooks and so on. The information in digital publication resource is normally

authoritative and useful. A digital encyclopedia is a kind of digital publishing resource holding a set of the important concepts from either all branches of knowledge or a particular branch of knowledge. Each concept is an entry in the encyclopedia. A domain-specific encyclopedia usually contains the main concepts in the domain. For example, the encyclopedia of the historical domain contains the major concepts related to history which include historical figures, historical events, and so on. These concepts are also mentioned in other more general texts as paragraphs or sections in e-books, digital magazines, digital newspapers, and so on. Generally, concept is an effective unit for cognition and learning. As a result, it is useful to reorganize the domain knowledge from the digital publishing resources by concepts. When users learn a concept, the related ones can be recommended to them for effectively learning. Recommender system has attracted increased interest nowadays because searching for relevant learning resources is a pivotal activity in TEL (Technology Enhanced Learning) [1,2]. Many recommender systems rely on users' preferences or history data [3–5]. It will not work for the circumstance such as cold start, scarcity of history data or preferences data. Semantic information is a possible method to overcome the problem under such circumstance. Semantic relatedness is computed by encyclopedias and then used for recommendation in the reference [6]. However, the computation depends totally on the labels and the explanation extracted from the encyclopedias. The content quality of the encyclopedia will largely impact the effect of recommendation.

Distributed words representations to be used in the paper were firstly proposed in [7]. It represents words as continuous vectors and the similar words are close in the vector space. Neural network can be used to learn the word vector and language model [8]. Learning semantic representation with a skip-gram model [9,10] was recently introduced by Mikolov *et al.* The model can learn the word vectors with simple neural network architecture so that it can be trained on a large amount of text in a short time. Word vectors can then be used to improve or simplify some applications [11,12].

A recommender system based on skip-gram model is presented in this paper without considering history data or preferences data of users when they learn the knowledge organized in concepts for a specific domain. Firstly, concepts of a domain are extracted from the domain-specific encyclopedia. Information in digital publishing resources are then reorganized and associated with these concepts. Secondly, skip-gram model is used to generate concept vectors by which the semantic relatedness is computed among the concepts. Therefore, users can learn the domain-specific knowledge organized by concepts and the related concepts can be recommended to them when they learn one of them. A few experiments have been conducted to validate the effectiveness of the method.

In the next section, we will describe the task in detail. We introduce the skip-gram model in the Section 3. In the Section 4, a new method is proposed to reorganize the information in the digital publishing resources and to implement the recommender system based on semantic relatedness. The Section 5 describes the results of the experiment. The conclusions are presented in the Section 6.

2. Problem Domain

Let $M = 0$ be the knowledge contained in a typical encyclopedia, where (x_i, y_i) is the set of labels and explanations in plain text of all concepts and A is other information such as figures and pictures. Let μ be the concepts mentioned in the set $H(y_i)$, where x_j represents a concept. Each concept o_i contains a label x_i , which would be the name of the concept, and an explanation y_i , which would be a

short piece of text which describes the concept. As a result, all labels will form a set $X = \{x_i, i = 1, \dots\}$. The information of the concept o_i can be extended by extracting information from the digital publishing resources and associating the information with the concept. The information for the concept o_i after extending can be represented as $K(o_i) = \{x_i, y_i, S_i, B_i, T_i\}$, where x_i, y_i, S_i, B_i, T_i is the label, explanation, sentence groups related to the concept, books related to the concepts, and other triples describing the concept. It can be seen that $K(o_i)$ reorganize the information from digital publishing resources in detail for the concept o_i . Users can learn the knowledge about the concept o_i through the information supplied by $K(o_i)$. To improve the efficiency of learning, other concepts should be recommended when users learn a certain concept. Let $f(x_i, x_j)$ be the semantic relatedness between the concepts o_i and o_j , where x_i and x_j are the vectors of the concepts o_i and o_j respectively. For the concept o_i , we should find $O'_i = \{o'_1, \dots, o'_n\} \subset O$, where $f(x'_j, x_i) \geq f(x_j, x_i)$, $o'_j \in O'_i, o_j \in O \setminus O'_i$, x'_j is the vector of the concept o'_j . The concepts in O'_i are more semantically related with o_i and the information they contain will help users to learn or understand o_i more effectively.

3. Word Distributed Representations by Skip-Gram Model

Skip-gram model is proven to be an efficient model to learn word distributed representations [9,10]. It can learn vector representations of words from large amounts of unstructured text data. The similar words are close in the vector space. The architecture of the skip-gram model is as the Figure 1.

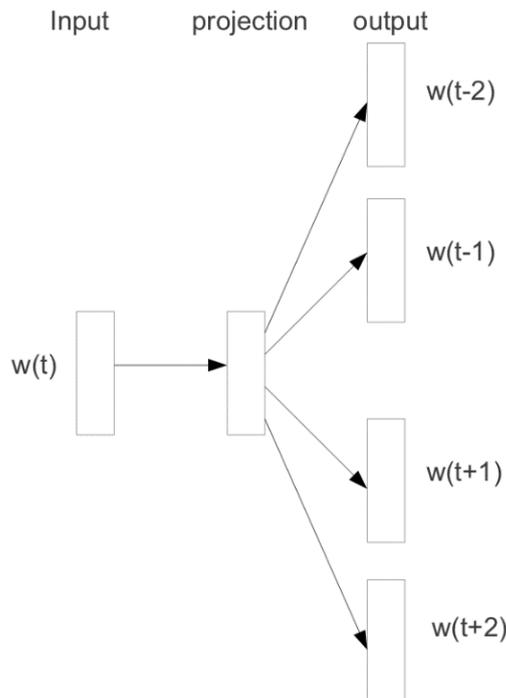


Figure 1. The architecture of the skip-gram model [9].

Given the training words sequence w_1, w_2, \dots , skip-gram model is to maximize the average log probability:

$$\frac{1}{T} \sum_{t=1}^T \left[\sum_{j=-k}^k \log p(w_{t+j} | w_t) \right] \tag{1}$$

where $j \neq 0$, and k is the size of the training window whose center is the word w_t . Larger k can lead to a higher accuracy due to more training examples considered. However, it will need more training time consequently.

Hierarchical softmax [13] is a computationally efficient approximation of the full softmax. The output layer will be represented by a binary tree with V words as leaf nodes, V is the number of words in the vocabulary. Each word w can be reached by a path from the root to the leaf node. Let $L(w)$ be the length of the path, $n(w, j)$ be the j th node on this path, $ch(n)$ be an arbitrary fixed child of the inner node n , $s(x)$ be 1 if x is true and -1 otherwise. Then $p(w | w_t)$ is defined by the formula:

$$p(w | w_t) = \prod_{j=1}^{L(w)-1} \sigma(s(n(w, j+1) = ch(n(w, j))) \cdot v_{n(w, j)}^T v_{w_t}) \tag{2}$$

where $\sigma(x) = 1 / (1 + \exp(-x))$, v_w is the representation of the word w and v_n is the representation of the inner node n .

A subsampling approach is used to counter the imbalance between the rare and frequent words. The probability to discard a word w_i in the training set is computed by the formula:

$$p(w_i) = 1 - \sqrt{\frac{t}{g(w_i)}} \tag{3}$$

where t is a specified threshold and $g(w_i)$ is the frequency of the word w_i . The method can accelerate learning and significantly improve the accuracy of the learned vectors of the rare words, which can be seen in [10].

The skip-gram model is trained using stochastic gradient descent algorithm by back propagation rule. The more information about the skip-gram model is described in the reference [9,10].

4. Recommender System Based on Semantic Relatedness

The knowledge in digital publishing resources is organized usually in chapters or sections. If the knowledge in a domain can be reorganized with concepts and the concepts can be recommended by their semantic relatedness, it will be greatly helpful and easy for reading and learning. Therefore, we implement the recommender system by reorganizing the domain knowledge in the format of concepts and computing the semantic relatedness among them. The domain-specific encyclopedia is an important resource for the process. The encyclopedias are selected from the digital publishing resources and concepts $i \neq j$ in the domain are extracted from them. The information of each concept o_i extracted from the encyclopedia consists of a label x_i and an explanation y_i . It is easy to get the label set $X = \{x_i, i = 1, \dots$ which contains the name of the concepts in the domain by traversing every concept. Sentence group set S_i can be extracted from the digital publishing resource for the concept o_i by the method we proposed in [14]. Then the books B_i from which the sentence group $s \in S_i$ is obtained are easily associated with the concept o_i . It is used to navigate the users to the original books when they learn the concept. The triples T_i which describe the properties and values of the concept o_i are extracted

from the text of digital publication resources. Some related methods or techniques might be found in the references [15,16]. As a result, the information of the concept o_i can be extended to $K(o_i) = \{x_i, y_i, S_i, B_i, T_i\}$.

The procedure to compute the semantic relatedness of the concepts is shown in the Figure 2. The label set $X = \{x_i, i = 1, \dots\}$ is firstly added to the dictionary of the word segmentation. The goal of this step is to ensure the label of the concept can be analyzed by word segmentation correctly. Text data is then extracted from the domain-specific digital publishing resources, such as encyclopedia, e-books, digital newspapers and so on. The text data is segmented by the word segmentation to form a document T which will be used to train the model. A vocabulary is built by counting the occurrence of each $x_i \in X$ in T which will be represented by $(x_i, \phi(x_i))$, where $\phi(x_i)$ is the occurrence of the concept $x_i \in X$ in T. The vocabulary is used to build a binary Huffman tree by the value of $\phi(x_i)$. The concept with larger value of $\phi(x_i)$ will have shorter unique binary codes in the binary tree. Then a skip-gram model is trained by the data T. When it is finished, each concept is represented by a vector with the dimension d . For each pair of concept vectors x and y , we can compute the semantic relatedness by the formula:

$$f(x, y) = \begin{cases} g(x, y) & \text{if } g(x, y) \geq \varepsilon \\ 0 & \text{if } g(x, y) < \varepsilon \end{cases} \quad (4)$$

where $g(x, y)$ is the cosine similarity of the two concept vectors x and y . The parameter $\varepsilon \in [0, 1]$ is used to control the minimal value of semantic relatedness between the two concepts.

$$g(x, y) = \frac{\sum_{i=1}^d (x_i \times y_i)}{\sqrt{\sum_{i=1}^d (x_i^2)} \times \sqrt{\sum_{i=1}^d (y_i^2)}} \quad (5)$$

A matrix M is created for the concepts by computing the semantic relatedness among the concept vectors. The size of the matrix is $n \times n$. The value of M_{ij} is set to $f(x_i, x_j)$ where x_i and x_j are the vectors of the concepts o_i and o_j respectively. The element in the diagonal of the matrix is set to zero. When the matrix is created, we can obtain the semantically related concepts easily. For the concept o_i , we get the i th row of the matrix M and put the n values in a list and resort the list by descending order. Get the concepts which are corresponding to the top m values in the list. These concepts form a set $O'_i = \{o'_1, \dots, o'_m\}$. The concepts in O'_i are more semantically related with o_i and the information they contain will help users to learn or understand o_i more effectively.

The main process of the recommender system is summarized as following steps. Most computation in each step is automatic in the system with the computation results checked by people.

Step 1: Select domain-specific encyclopedia and extract concepts $O = \{o_1, \dots, o_n\}$ from the encyclopedia by regular expression. Each concept o_i is consisted of a label x_i and an explanation y_i .

Step 2: For each concept o_i , associate sentence groups, books, and triples describing the concept to form $K(o_i) = \{x_i, y_i, S_i, B_i, T_i\}$ which is described above.

Step 3: Add the labels $X = \{x_i, i = 1, \dots, x_n\}$ to the dictionary of the word segmentation.

- Step 4:** Extract text data from domain-specific digital publishing resources and segment the text data to generate a document T . The document contains the concepts and their context.
- Step 5:** Build a vocabulary by counting the occurrence of each $x_i \in X$ in T .
- Step 6:** Create binary tree by the vocabulary and train the skip-gram model by the document T . After training finishes, get vector for each concept from the model.
- Step 7:** For each pair of concepts, compute the semantic relatedness by the Equation (4). We get a $n \times n$ matrix M in which M_{ij} is set to the semantic relatedness of the concepts o_i and o_j .
- Step 8:** For each concept o_i , generate $O'_i = \{o'_1, \dots, o'_n\}$. The concepts in the set O'_i are more semantically related with o_i .
- Step 9:** When users learn a concept o_i , provide the users with the information $K(o_i) = \{x_i, y_i, S_i, B_i, T_i\}$. At the same time, recommend the concepts in the set $O'_i = \{o'_1, \dots, o'_n\}$ to them.

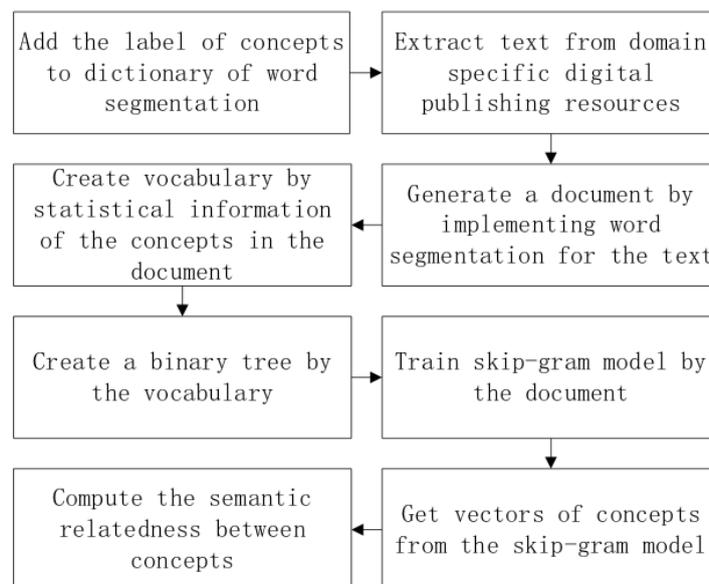


Figure 2. The process to compute the semantic relatedness of the concepts.

5. Experiments

The goal of the experiment is to investigate whether the method proposed is effective to recommend the related concepts which are organized by the data of digital publishing resources. In our experiment, concepts are extracted from three books of encyclopedias in the domain of history by setting the proper regular expression. The titles of the books are Encyclopedia of China, Chinese history I, II, III [17]. These books contain the important concepts of Chinese history. In these books, 2310 concepts are extracted and selected with their labels and explanations. About 20,000 e-books related with history knowledge are used to extract sentence groups and triples which are then associated with the concepts. When a sentence group is associated to a concept, the e-book from which the sentence group is extracted is also connected with the concept so that users can review the original books for the more information easily. To compute the semantic relatedness among the concepts, we select 27 history related books from the e-books and use the text after word segmenting for training. Some books of them are General History

of China, History of Ancient China, Chinese History and Culture, *et al.* The vocabulary is generated by the concept labels and the text. The vocabulary is used to create binary tree in the output layer. In the experiment, the dimension of the concept vector is set to 50. The context window size is set to 20 and the subsampling threshold is set to 1.0×10^{-3} . When training is finished, vectors of all concepts are generated.

We create a relatedness matrix for each pair of concepts by computing the semantic similarity among the concept vectors. The size of the matrix is 2310 rows and 2310 columns. The value of each item in the matrix is computed by the Equation (4). Since $f(x, y) = f(y, x)$, we need not compute both but one of them. The value of $f(x, x)$ is set to zero. It means that the diagonal of the matrix will be zero. After the relatedness matrix is generated, we can get the semantic relatedness between any pair of concepts by the concept's row and column index in the matrix. For example, we can get the row index of the concept “秦始皇” (Qinshihuang) in the matrix and get all values in this row. Qinshihuang is a famous historical personage who was the first king in Qin Dynasty. The values represent the semantic relatedness between Qinshihuang and the other concepts. The concepts are then sorted by the value of semantic relatedness with Qinshihuang in descending order. When we set the parameter m to 30, the top 30 concepts can be recommended to users when they learn the concept Qinshihuang. Table 1 lists some of the concepts recommended. The column Concept represents the concepts which are recommended and the column Relatedness shows the semantic relatedness. In the table, “焚书坑儒” (burn books and bury alive Confucian scholars) and “陈胜吴广起义” (the uprising of Chensheng and Wuguang) are historical events happened in “秦朝” (Qin Dynasty). “半两” is the coin name which was used in Qin and early Han Dynasty. “相邦” is a government official title in Qin Dynasty. “吕氏春秋” is an ancient Chinese chronicle which is collected and arranged by “吕不韦” and hangers-on. “吕不韦”, “李斯” and “赵高” are the Prime Minister of Qin Dynasty. “黔首” is a term used in ancient China which represents the common people. “蒙恬” is a military officer in Qin Dynasty. “郡县制” is a system of local administration which took shape during the Spring and Autumn Period and the Qin Dynasty. “秦简” is bamboo book of Qin Dynasty. “云梦秦律” is the law of Qin Dynasty. “秦二世胡亥” is the secondary king in Qin Dynasty. “灵渠” is a canal created in Qin Dynasty. “秦郡” is the administrative planning system of the Qin Dynasty. “封禅” is a grand ceremony of worship of heaven on mountain top to pray and say thanks for peace and prosperity in ancient China. It can be seen from the table that the concepts recommended have closely semantic relationship with Qinshihuang. Therefore, the user can review the detail information of the related concepts to help him understand well the concept of Qinshihuang. The relatedness is represented by the value in the interval $[0, 1]$, which will be useful for us to adjust the view according to the relatedness to determine the number of concepts displayed. It is helpful for users to learn or understand the concepts more effectively in e-learning environments.

Since the semantic relatedness in the recommender system is computed by the text of digital publishing resources which is normally written in different languages holding a relatively complete and authoritative collection of concepts and texts in a specific domain, the proposed method can be used in the different language environments and can cover almost all important concepts in a domain. This is not the case for WordNet-based method [18], ESA [19] or WikiRelate! [20]. WordNet based method can only be used for English. ESA and WikiRelate! need Wikipedia for computation. However, there may be not enough entries of Wikipedia in some languages nowadays. The method proposed in the reference [6] is based on the explicit relation and implicit relation computed by encyclopedia of digital

publishing resources. However, the computation depends totally on the labels and explanation extracted from the encyclopedias. The content quality of the encyclopedia will largely impact the effect of recommendation. The method proposed in this paper will consider not only encyclopedias but also other digital publishing resources. It can compute the relationship from various digital publishing resources in a more comprehensive perspective and reduce the impact of the encyclopedias. Furthermore, this method need not consider history data or users' preferences data for recommendation in a specific domain. When a new user uses the e-learning system, the most popular concepts can be displayed for the user. When the user click and learn one of them, the related concepts computed by the method can then be organized for recommendation. After the system collects enough data, the recommendation can then be refined with the personalized data. As a result, it can work in the circumstances such as cold start, scarcity of history data or preferences data.

Table 1. Concepts recommended for the concept Qinshihuang.

Concept	Relatedness	Concept	Relatedness	Concept	Relatedness
焚书坑儒	0.874802411	黔首	0.873911211	李斯	0.855913035
秦朝	0.838465314	蒙恬	0.826526496	秦二世胡亥	0.807738280
陈胜吴广起义	0.800022638	赵高	0.773887009	灵渠	0.767677992
半两	0.767172869	郡县制	0.734288249	吕不韦	0.672464491
相邦	0.670248099	秦简	0.669399414	秦郡	0.666014029
吕氏春秋	0.650068964	云梦秦律	0.640233037	封禅	0.620420199
...

The method proposed in the reference [6] is also to implement a semantic recommender system by the semantic relatedness of concepts. It will be used as baseline in this experiment.

Kendall concordance coefficient (Kendall tau) is widely used to measure the association between two measured quantities. Let $(x_1, y_1), (x_2, y_2), \dots$ be a set of observations of the joint random variables X and Y respectively. Any pair of observations (x_i, y_i) and (x_j, y_j) are said to be concordant if $(x_i - x_j)(y_i - y_j) > 0$. They are said to be discordant if $(x_i - x_j)(y_i - y_j) < 0$. If $(x_i - x_j)(y_i - y_j) = 0$, the pair is neither concordant nor discordant. The Kendall tau is defined as:

$$\tau = \frac{2(c - d)}{n(n - 1)} \tag{6}$$

In the Equation (6), c and d are number of concordant pairs and number of discordant pairs respectively. It is used in this section to compare the data produced by the algorithms and the data prepared by human participants.

One hundred entities are selected randomly from the encyclopedias [17]. Four people were invited to assign values of semantic relatedness from the concept Qinshihuang to the entity and they did the job separately. Four groups of data produced by them are represented as P1, P2, P3, P4 and the data produced by the method in the reference [6] is represented as X. The data generated by the algorithm proposed in this paper is represented as Y. The parameter ϵ is set to 0.1. Kendall tau is then computed with data between any two groups. The result is listed in the Table 2. The field Avg. is the average value of five pairs. For example, the average Kendall tau is calculated by five pairs between X and P1, X and P2, X and P3, X and P4, X and Y on the first line.

Table 2. Kendall tau between six groups of data.

Group	X	P1	P2	P3	P4	Y	Avg.
X		0.38	0.31	0.22	0.41	0.30	0.32
P1	0.38		0.49	0.34	0.47	0.40	0.42
P2	0.31	0.49		0.35	0.45	0.39	0.40
P3	0.22	0.34	0.35		0.35	0.29	0.31
P4	0.41	0.47	0.45	0.35		0.36	0.41
Y	0.30	0.40	0.39	0.29	0.36		0.35

From the Table 2 it can be seen that Kendall tau of the groups P1, P2, P3, P4 is normally less than 0.5. This is mainly for two reasons. One is that the knowledge each person holds is different from another's knowledge; it is hard for people to assign the semantic relatedness consistently. The other is that there are many pairs with the same value in the data, especially: there are many zero values which are assigned by people for some relation. This means a person can hardly assign a differentiable value for each relation. However, we need the differentiable value for each pair of relations so that it can be measured and used for recommendation effectively. As a result, it is necessary and important to use the method to assign the differentiable semantic relatedness objectively and consistently for a recommender system.

The first line shows that the average Kendall tau between X and the other five groups of data is 0.32. The last line shows that the average Kendall tau between Y and the other five groups of data is 0.35. This means that the method proposed in this paper may work better than the one of the reference [6] in this circumstance. The main reason is that the method proposed in the reference [6] depends mainly on the explicit relation between the concept's labels and explanation extracted from the encyclopedias. The implicit relation is computed by the path in the explicit relation graph. The relation is largely determined by the information in the path. However, the relation can be learned in this method from the text where the concept can be considered in more dimensions by the context. It can describe the relation in more comprehensive view.

6. Conclusions

A lot of useful and authoritative information is contained in digital publishing resources. The information is normally organized in sections and paragraphs. To reorganize the resources and recommend the related concepts is very helpful for e-learning. A recommender system is presented for e-learning in this paper based on the semantic relatedness of concepts computed by the text of digital publishing resources. Semantic relatedness is computed based on skip-gram model for the concepts extracted from the domain-specific encyclopedia. The related concepts and associated information will be recommended to users for their reading or learning. The recommender system does not need users' history data or preferences data. It takes into account not only encyclopedias but also other digital publishing resources for semantic relatedness computing to reduce the impact of the quality of encyclopedias. The proposed method can be used in the different language environments and can cover almost all important concepts in a domain. The experiment shows its potential usability for e-learning in a specific domain. The next steps include combining the relatedness with the distance of the concepts in the ontology built for the domain of Chinese history, and refining the recommendation with the personalized data collected by the system.

Acknowledgments

The work was supported by China Postdoctoral Science Foundation (2013M540789) and Beijing Postdoctoral Research Foundation (2013).

Author Contributions

Zhi Tang and Jianbo Xu designed research; Mao Ye and Lifeng Jin performed research and analyzed the data; Mao Ye wrote the paper. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. 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 Trans. Learn. Technol.* **2012**, *5*, 318–335.
2. Manouselis, N.; Drachsler, H.; Verbert, K.; Duval, E. *Recommender Systems for Learning*; Springer: Heidelberg, Germany, 2013; pp. 1–20.
3. Yolanda, B.F.; Jose, J.P.A.; Alberto, G.S.; Manuel, R.-C.; Martín, L.-N.; Jorge, G.-D.; Ana, F.-V.; Rebeca, P.D.-R.; Jesús, B.-M. A Flexible Semantic Inference Methodology to Reason about User Preferences in Knowledge-based Recommender Systems. *Knowl.-Based Syst.* **2008**, *21*, 305–320.
4. Carrer-Neto, W.; Hernandez-Alcaraz, M.L.; Valencia-Garcia, R.; García-Sánchez, F. Social Knowledge-based Recommender System, Application to the Movies Domain. *Expert Syst. Appl.* **2012**, *39*, 10990–11000.
5. Bobadilla, J.; Ortega, F.; Hernando, A.; Gutiérrez, A. Recommender Systems Survey. *Knowl.-Based Syst.* **2013**, *46*, 109–132.
6. Ye, M.; Jin, L.F.; Tang, Z.; Xu, J. A Semantic Recommender System for Learning Based on Encyclopedia of Digital Publication. In Proceedings of the International Conference, HCI International 2014, Heraklion, Greece, 22–27 June 2014; Springer: Cham, Switzerland, 2014; pp. 189–194.
7. Rumelhar, D.E.; Hinton, G.E.; Williams, R.J. Learning Representations by Back-propagating Errors. *Nature* **1986**, *323*, 533–536.
8. Bengio, Y.; Ducharme, R.; Vincent, P. A Neural Probabilistic Language Model. *J. Mach. Learn. Res.* **2003**, *3*, 1137–1155.
9. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. In Proceedings of the ICLR Workshop, Scottsdale, AZ, USA, 2–4 May 2013.
10. Mikolov, T.; Sutskever, I.; Chen, K.; Dean, J. Distributed Representations of Words and Phrases and Their Compositionality. In Proceedings of the NIPS, Lake Tahoe, NV, USA, 5–10 December 2013.
11. Mikolov, T.; Le, Q.V.; Sutskever, I. Exploiting Similarities among Language for Machine Translation. 2013, ArXiv:1309.4168v1.

12. Turian, J.; Ratinov, L.; Bengio, Y. Word Representations: A Simple and General Method for Semi-Supervised Learning. In Proceedings of the Association for Computational Linguistics, Uppsala, Sweden, 11–16 July 2010.
13. Morin, F.; Bengio, Y. Hierarchical Probabilistic Neural Network Language Model. In Proceedings of the International Workshop on Artificial Intelligence and Statistics, Bridgetown, Barbados, 6–8 January 2005; pp. 246–252.
14. Ye, M.; Jin, L.F.; Tang, Z.; Xu, J. Sentences Extraction from Digital Publication for Domain-specific Knowledge Service. In Proceedings of the International Conference, HCI International 2014, Heraklion, Greece, 22–27 June 2014; Springer: Cham, Switzerland, 2014; pp. 274–279.
15. Zouaq, A.; Nkambou, R. Evaluating the Generation of Domain Ontologies in the Knowledge Puzzle Project. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1559–1572.
16. Gaeta, M.; Orciuoli, F.; Paolozzi, S.; Salerno, S. Ontology Extraction for Knowledge Reuse: The E-Learning Perspective. *IEEE Trans. Syst. Man Cybern. Part A: Syst. Hum.* **2011**, *41*, 798–809.
17. Encyclopedia of China, Chinese History I,II,III. Available online: <http://www.apabi.com/apabi> (accessed on 10 February 2014).
18. Budanitsky, A.; Hirst, G. Semantic Distance in WordNet: An Experimental, Application-oriented Evaluation of Five Measures. In Proceedings of the Workshop on WordNet and Other Lexical Resources, Pittsburgh, PA, USA, 3–4 June 2001.
19. Gabrilovich, E.; Markovitch, S. Computing Semantic Relatedness Using Wikipedia-based Explicit Semantic Analysis. In Proceedings of the IJCAI, Hyderabad, India, 6–12 January 2007; pp. 1606–1611.
20. Strube, M.; Ponzetto, S.P. WikiRelate! Computing Semantic Relatedness Using Wikipedia. In Proceedings of the AAI, New York, NY, USA, 4–9 June 2006; pp. 1419–1424.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).