Discovery and Recommendation of First-Hand Learning Resources Based on Public Opinion Cluster Analysis

Haiyun Li(✉), Xuebo Zhang, Junhui Wang
South China Normal University, Guangzhou, China
`harwine@163.com`

Abstract—This paper explores the personalized approach of the public opinion cluster analysis for learning resources based on the server-side predetermined analysis, in order to introduce the personalized learning resource recommender into the traditional online instruction. In allusion to further validation on its implementation, the fuzzy aggregation of learning resources is mined up based on the proposed WRTC algorithm. The personalized learning resource recommender mechanism is then described. In the end, the common evaluation parameters in the personalized recommender model are applied in the evaluation on the system performance. The experiment is carried out with learner's access data online to validate whether the algorithm and the model indicators are effective for the purpose of improving the precision and coverage of learning resources.

Keywords—personalized learning recommender, public opinion, text clustering

1 Introduction

Online Learning System (OLS) has aroused people's wide concern with the dramatic development of information technology, however, the disadvantages of the traditional online instruction model have gradually been exposed to the view of people in the application. The monotonous learning resources are no longer satisfying the learners' increasing demands, especially when the differentials between individuals are hushed up, this contradiction becomes more prominent, as stated in [1]. While the personalized instruction system based on the online public opinion emerges formally to cater to the individual wishes of more learners at the right moment, which is designed to best meet all types of learners' demands in view of their different learning interests, cultural levels and knowledge degrees, and maximize their freedom and initiative in learning process, as stated in [2]. On this basis, the different knowledge systems and learning demands will be targeted to recommend learning resources to promote the effective improvement of students' academic records.

The constructivist theory, the humanism theory and the multiple intelligence theories, as the theoretical foundations for personalized learning, all attribute to web-based instruction models "as stated in [3]", which apply the cluster analysis technology to

extract the most valuable and useful information from massive data as required for demanders. Whereas the cluster analysis technology based on public opinion can play a better cluster analysis technology which furnishes the individualized learners with targeted learning resources by data parser and extractor, "as stated in [4]". It has therefore gradually caught more and more people's eye.

2 Cluster analysis technology

2.1 Overview

It refers to the process where the physical objects or the abstract objects are classified into multiple groups according to different types and features. It is worth noting that data preprocessing, the definition of distance function, the aggregations and the evaluation on clustering results are 4 important procedures in a typical cluster process, "as stated in [5]".

Table 1. Traditional clustering algorithm

Traditional clustering algorithm	Classical algorithm
Level method	BIRCH, CURE, ROCK, CHEMALOEN
Division method	K-MEANS, K-MEDOIDS, CLARA, CLARANS
The density based method	DBSCAN, OPTICS
The grid based method	STING, CLQUE

Preprocess data. Aiming at the object with high feature dimension as it is originally, it may achieve the goal of feature standardization and dimension reduction by means of feature selection and extraction. First, the purpose of feature selection is to reduce the spatial dimension in the principle of excluding the features that affect the consequences. Feature extraction is mainly applied in the case of high original feature dimension to transform and map the original general features into new features, thus making a presentation on the original sample with low-dimensional space.

Define the distance function. Cluster analysis technology should be implemented based on the inter-object similarity measured by the distance function. On this basis, it is of great importance to measure the similarity of different objects across one space.

Aggregation. The cluster algorithm is applied to classify the objects in question according to their different characteristics.

Evaluation on aggregation results. The clustering results are estimated to be as an important basis for determine whether the clustering quality is good. In general, the indicators for aggregation evaluation are cluster distance and intra-class coupling degree, while the aggregation effect is positively correlated to the growth of the above two indicators.

The traditional clustering algorithm is mainly divided into the following four types, see Table 1.

2.2 Cluster analysis approach for learning resources

In general, the learning resources accessed by a user must have certain relevance by which the learning resources are aggregated by analyzing the frequency of user accesses to learning resources. "As stated in [6]". In the cluster analysis algorithm for learning resources, different class attributes are set for different learning resources to make a representation on the properties of their respective resources.

Definition 1. Similarity of Learning Resources: the similarity Rij is used to represent the percentage of user sessions of the learning resources i and j, that is, the ratio of user sessions which includes both the learning resources i and j in Tij to the total user sessions in T. In general, the closer to 1 the Rij, the higher the similarity between two learning resources is. If Rij = 1, there is no difference between the two resources.

Definition 2. Similarity Matrix of Learning Resources: Rij represents the similarity between the elements in the matrix; the serial number of the learning resource is used as the row or the column in the similar matrix S. Rij definition analysis shows that the matrix S is a fuzzy similarity in line with the symmetry and reflexivity.

Definition 3. Transitive closure for learning resources: including the transitive matrix of S with minimum similarity.

In the aggregation of learning resources based on transitive closure, the similarity matrix S for learning resources is first constructed based on the similarity calculation of two given learning resources. The transitive closure T for similarity matrix of the learning resources is then calculated, where the elements are sorted to obtain the preferred sequence of cluster intercept set γ. The precision of the classification is proportional to value γ. After determining the value γ, the aggregation results are output. The specific process is shown in Fig. 1 below:
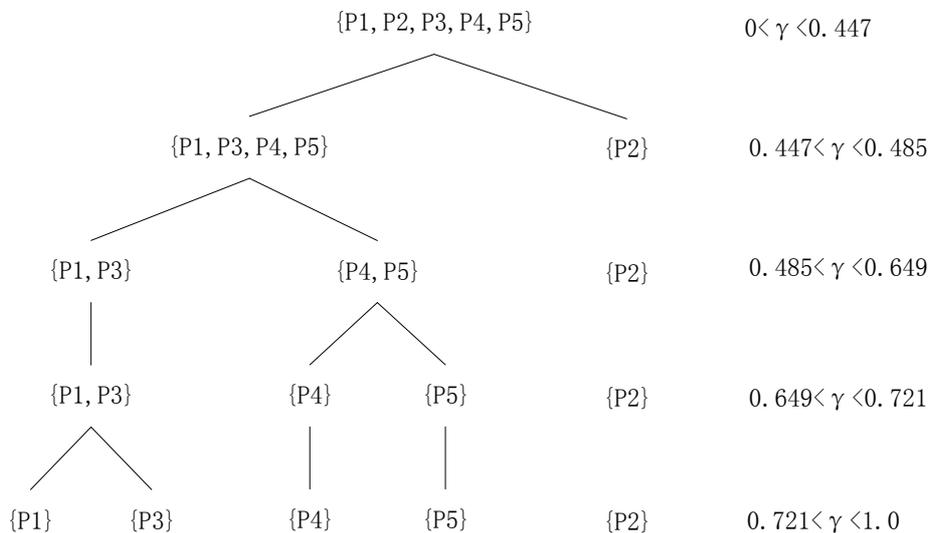

Fig. 1. Aggregation results

3 Discovery and recommendation of learning resources based on public opinions

3.1 Personalized recommender algorithm

The WRTC (Recommendation Algorithmbasedon Weighted Association Rules and Transitive Closure Cluster) recommender algorithm is proposed based on weighted association rules improved for the transitive closure fuzzy cluster algorithm and the apriori algorithm, "as stated in [7]", the specific procedure is given as below:

Input: server-side public opinion data

Output: recommended learning resources, and its precision and coverage;

Step 1: The aggregation of the trained datasets is realized by the transitive closure fuzzy cluster algorithm.

Step 2: The trained data set is mined using the weighted association rules;

Step 3: It mainly calculates the recommender value of learning resources, and is usually a comprehensive calculation using the resource clustering recommender values and the confidence of the association rules.

Step 4: The recommended value is available and analyzed to recommend n learning resources with higher recommended value for user.

In the WRTC algorithm, the first step is to make a measure for the similarity of the current user sessions, based on which to calculate the recommended value; the weighted confidence of association rules and the recommended value are also used simultaneously to capture the final recommendable resources while current user is matched with the weighted association rules. It should be noted that the match rate between the current user session and the weighted association rule refers to a higher similarity rule instead of a precision rule.

3.2 Experimental verification and evaluation criteria

The coverage and precision are two key indicators for measuring whether the recommender algorithm is good, "as stated in [8]". The recommender precision means the ratio of correct recommendations to the total, where the right precision refers to the contents of the learning resources accessed by learners in the recommended resources. Let R (p) denote the set of accessed resources the system recommends to user, where the user-assessed resources are p; T(p) represents the all resources accessed by user in this session after p; the recommended precision is calculated by the formula (1) :

$$\Pr ecision = \frac{T(p) \cap R(p)}{R(p)} \tag{1}$$

The coverage is the ratio of recommended resources to the total resources accessed after the current user session. It is calculated as below:

$$Coverage = \frac{T(p) \cap R(p)}{T(p)} \tag{1}$$

3.3   Experimental procedure

Input: server-side public opinion data in a network academy, a total 1003 entries
Output: recommended results for acquiring the learning resources.

Procedure: Step 1: As shown in Fig. 2 below, the public opinion data of the network academy is preprocessed in the process, and 274 entries of useful information are finally available in total;
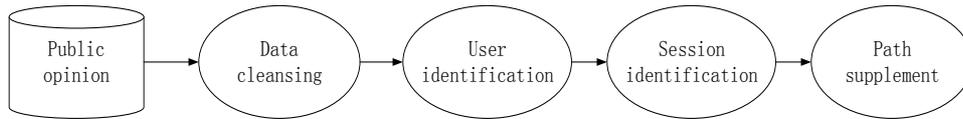

Fig. 2. Data preprocessing process

Step 2: The resulting user sessions are classified as follows: the class 1 is a training dataset including 18 user sessions; the class 2 is a test dataset including 10 user sessions, which is then revised: these learning resources with prices are arranged according to the degree of interests in descending order, and the first n learning resources are set in the Slide Window.

The digital order of accessing learning resources is calculated by following formula (2):

$$Fresh(p_i) = \frac{i}{|w|} \qquad i = 1, 2, \quad |w| \tag{2}$$

Where, the first resource to be accessed is represented by 1, the size of slide window is by w, and i denotes the position from where the learning resource can be accessed in the slide window. The last resource has maximum value Fresh. Only Website weights must be normalized can it be ensured to effectively respond to the impact on the access order of learning resources, "as stated in [9]", the specific formula is given as below:

$$W_{normalized}(p_i) = \frac{w(p_i)}{\sum_{j=1}^{n} w(p_i)} \tag{4}$$

Summary and analysis of the above program show that, Wnormalized and Fresh have equal impacts, i.e. at a same proportion. To ensure that the degree of interest in learning resources is increased when both Fresh and Wnormalized are higher, we usually express this interest degree as the average of the two "as stated in [10]", see formula (5) for specific calculation:

$$Interest(P_i) = \frac{2 \times Fresh(p_i) \times W_{normalized}(p_i)}{Fresh(p_i) + W_{normalized}(p_i)} \tag{5}$$

Step 3: Use the transitive closure cluster algorithm to aggregate the training data set for learning resources of 6 categories.

Step 4: Calculate the appropriate association rules in the training data set using the price association rule algorithm, and finally 231 weighted association rules are available.

Step 5: Incorporate the recommended value in the resource aggregation and the confidence level of the association rule to obtain the recommended values of learning resources;

Step 6: arrange learning resources obtained in descending order by recommended values, and the first n learning resources are recommended to the front users;

3.4 Experimental results

The cluster algorithm and the association rule mining algorithm are all widely applied in the field of personalized recommender. The transitive closure cluster algorithm and association rule algorithm and Apriori can be selected as the reference objects to evaluate the experimental object. Different learning resources recommended (1-7) may be given, respectively using the transfer closure cluster algorithm, WRTC and Apriori as mentioned in this paper for recommender coverage and precision of learning resources with the results as shown in Tables 4.3 and 4.4 below. The study shows that the WRTC is significantly better than the other two algorithms when calculating the recommended coverage and precision of learning resources.

Table 2. WRTC algorithm and association rule algorithm, clustering algorithm to achieve web page referral coverage comparison

	1	2	3	4	5	6	7
Apriori	0.3751	0.4224	0.4583	0.4622	0.4802	0.5102	0.5610
Cluster	0.3513	0.4235	0.4775	0.4816	0.4961	0.5213	0.5721
WRTC	0.3892	0.4335	0.4878	0.5064	0.5187	0.5403	0.5887

Table 3. WRTC algorithm and association rule algorithm, clustering algorithm to achieve resource recommendation when the precision comparison

	1	2	3	4	5	6	7
Apriori	0.7400	0.6876	0.5808	0.5501	0.5061	0.4097	0.4598
Cluster	0.6802	0.6545	0.5678	0.5151	0.4878	0.4407	0.4287
WRTC	0.7802	0.7731	0.7087	0.6834	0.6378	0.5958	0.5830

4 Conclusion

This paper mainly describes the typical algorithm of public opinion mining, i.e. the cluster algorithm based on the transitive closure, which is implemented at the frequency of learning resources being accessed simultaneously in a user session. The learning resources are classified by the similarity, in order to achieve the goal of nar-

rowing the recommended scope. Beyond that, this paper also introduces and explores the personalized recommender algorithm based on fuzzy clustering algorithm for learning resources. In the end, the experiment is conducted to validate the model indicators and relevant algorithms for improving the coverage and precision of learning resources recommender with public opinion data from online access of a learner,

5 References

[1] Pragya, D. Kamal, K. B. (2015). E-Learning recommender system for a group of learners based on the unified learner profile approach. Expert systems, 32(2): 264-276. https://doi.org/10.1111/exsy.12061

[2] Outmane, B. Essaid, E. B. Mohamed, E. A. (2016). A Personalized E-Learning Based on Recommender System. International Journal of Learning and Teaching, 2(2): 99-103.

[3] Dianshuang, Wu. Jie, L. Guangquan, Z. (2015). A Fuzzy Tree Matching-Based Personalized E-Learning Recommender System. IEEE Transactions on Fuzzy Systems, 23(6): 2412-2426. https://doi.org/10.1109/TFUZZ.2015.2426201

[4] Baocheng, H. Guang, Y. (2015). Research and application of public opinion retrieval based on user behavior modelling. Neurocomputing, 167: 596-603. https://doi.org/10.1016/j.neucom.2015.04.029

[5] Ning, M. Yijun, L. (2014). SuperedgeRank algorithm and its application in identifying opinion leader of online public opinion supernetwork. Expert Systems with Applications, 41(4): 1357-1368. https://doi.org/10.1016/j.eswa.2013.08.033

[6] Antonio, G. Massimo, P. Marzia, B. (2017). On the Application of Text Clustering in Engineering Change Process. Procedia CIRP, 62: 187-192. https://doi.org/10.1016/j.procir.2016.06.019

[7] Jiaming, X. Bo, X. Peng, W. (2017). Self-Taught convolutional neural networks for short text clustering. Neural Networks, 88: 22-31. https://doi.org/10.1016/j.neunet.2016.12.008

[8] Laith, M. A. Ahamad, T. K. Mohammed, A. A. (2017). Text feature selection with a robust weight scheme and dynamic dimension reduction to text document clustering. Expert Systems with Applications, 84: 24-36. https://doi.org/10.1016/j.eswa.2017.05.002

[9] Duan, H.Y. (2016). Research on Collaboration in Innovative Methods of Manufacturing Innovation Chain. Revista Iberica de Sistemas e Tecnologias de Informacao, E11: 292-303.

[10] Baojun, M. Nan, Z. Guannan, L. (2016). Semantic search for public opinions on urban affairs: A probabilistic topic modeling-based approach. Information Processing & Management, 52(3), 430-445. https://doi.org/10.1016/j.ipm.2015.10.004

6 Authors

Haiyun Li, Xuebo Zhang, and **Junhui Wang** are with South China Normal University, Guangzhou, China.