Journal of Education and Praxis Research. Volume 6, Issue 1, (2020) 10-17
http://www.cayley-nielson.com/jepr

Applying Collective Intelligence Technology for Energy Economic Knowledge Database Recommendation System

Lisa Marcoux*
Mathematics Computer Science Department, Glendale Community College, Maricopa Community Colleges, 6000 West Olive Avenue, Glendale, Arizona 85302, USA
*Corresponding author
E-mail: *LIS2171314@maricopa.edu*

Abstract

This study has developed a data customize model to construct the energy economic knowledge database. Collective intelligence technology is used to do knowledge recommendation in the system. In this paper, Slope one algorithm was initially introduced into the energy economic knowledge recommendation and the feedback of the recommendation result will be used for self-learning rating. We propose a personalized recommendation method joins weighted slope one algorithm and clustering algorithm, which will improve the calculation accuracy and reduce the computational complexity. We demonstrate the usefulness of the method on a dataset. The method has been proved to be useful for obtaining a good recommendation for energy economic online teaching.

Keywords: Collective Intelligence; Knowledge Management; Online Teaching.

1. Introduction

Knowledge management system, according to the theory of knowledge management, completed classified storage and management of a large number of valuable programs, planning, results and experience. It promotes learning, sharing, training, and innovation. A good knowledge management system, from the point of the instrumental, it provides knowledge creation, review, publish, use, interact, share, push, evaluation, assessment, analysis, sorting and other specific features.

Network teaching by knowledge management system should make each user learn to finish the accumulation of knowledge and information through teaching network system [8]. The teacher should let the students learn through the network to accumulate new knowledge to help improve the teaching level.

In the future of online teaching, user learning is no longer repeatable process to copy the existing knowledge of their own minds, but continue to supplement and improve the existing knowledge combined with search such as internet data. Users can search information on the internet to find a lot of new knowledge, and share valuable knowledge to other users [9].

Collective intelligence is based on crowd behavior research, and do the personalized recommendations for each user, so that users can accurately information they need more quickly [5]. Today's successful recommendation engine does not require rigorous modeling of items or users, and does not require a description of machine understandable. Collective intelligence is the recommended method which has nothing to do with the field. At the same time, this method is open, you can share the experience to others, make users find potential interest preferences. The core of collective intelligence is the recommend function, and finishes this by exploring the user's behaviors and other users with similar behaviors.

ISSN 2378-9654 /
May, 2020

The evaluation of students' learning, not primarily to see how much prior knowledge they learned, but look the number of new discoveries, new insights, new materials and new contribution in the related disciplines of the learning process [10]. So the knowledge management system is important for online teaching.

2. Network data integration

Knowledge management system architecture should include the core database, knowledge management system, intelligent search engine. How effective the data was integrated into the knowledge database is the basis of the knowledge management system, especially the integration of network data to the knowledge management systems. And it provides users with a complete knowledge.

Network data can be viewed as a larger, more complex database, each site on the network is a data source, and each data source is heterogeneous. The data on the network is very complex, no specific model description, the data on each site designed independently. Thus, the data on the network is a non-fully structured data, which is also known as semi-structured data and unstructured data, such as web pages, user discussion and thoughts on the stock, the customer feedback e-mail or forum, etc. So the data integration process on the network is different from traditional database [7].

The challenge faced by knowledge management system is how to achieve correct data through different information platforms, how to integrate a large number of available data and transform them into information assets as soon as possible. Because of the source data and the format is different, so data integration is difficult. And all these require a powerful data integration technology. But as the early data sources mainly considering various relational databases, therefore, the integration is mainly pointed to relational database.

Data extraction, transform and loading solution contains three areas, first, 'selected': read out raw data from various web sites, it is a prerequisite to all the work. Followed by is 'conversion': to obtain data conversion pumping according to the pre-designed rules and the heterogeneous data formats here can be unified. Finally, the 'load': converting the data as planned incremental or all and import the data into the database.

Transform steps generally include the process called data cleaning; data cleaning is mainly directed against the source database for the occurrence of ambiguity, duplication, incomplete or logic corresponding data. Data quality analysis is required before cleaning operation; it will identify problems in the raw data. Data loading process could use loading tool or load the data by the SQL program and then loaded the results of the data extraction, transform into the target database.

3. Search optimization

Today, we can use search engines to find information of our interest. For example, if we want to see the weather forecast, look for a product price, and compare a product prices in different stores, we can use search engines. However, these ways of using search engines is passive, we must tell the search engine what we want to search. In other words, for energy economic knowledge online teaching users, including information on energy products information, national policy and the fundamentals of companies and other information, users have to search manually. And the shortage of this way of accessing information is that the user could not obtain latest update information in time.

Currently, the search engine does not fully meet the needs of users, it is because in many cases, the user is actually not clearly knowing their own needs, or express their demand is difficult by using simple keywords.

For energy economic knowledge network teaching, what the users concerned about most is the related information that impacts their learning subjects. The subject impacted by many factors, such as the national policy, the fundamentals of companies and other factors. With the development of information technology, users wish to get new information in time, such as the latest price of a product and all the latest information which can affect the volatility of a product [6]. Users wish to customize the information according to their interests, and the system should capture the information from the massive real-time network information, and organize the information, turn it to be standard, orderly information resources. Finally, the system should do the real-time push according to the users customize keywords.

Push technology as a major way of accessing information, it could push the information of interest to the user automatically, and help users explore the valuable information efficiently. In the energy economic information push application, the user can take the initiative to customize their own keywords, such as customize "deposit rates ", "oil prices", "economic policy" and other related keywords. When the system retrieves the latest information and research reports associated with the keywords, it can push to the users.

We could construct a user customize model to present the data as follows:

```
< Account>
<User>
<Name>Bill Buckram</Name>
</User>
<Customize>
<Keyword>Oil prices</Keyword>
< Keyword > Energy economic</Keyword>
< Keyword >Economic policy</Keyword>
</Customize>
</Account>
```

From above, we know that the user customizes "oil prices", "economic policy" and "financial" keywords. He is concerned with the keywords of the energy financial policy. So we can analyze the behavior of user groups through analyzing them customize keywords, and analyze the users who have the same interest, this will provide a great value for customer relationship management application.

4. Cluster analysis

After Keywords customized similarity analysis, we can do the cluster analysis process. Cluster analysis is the task of assigning a set of objects into groups so that the objects in the same cluster are more similar to each other than to those in other clusters. Teachers who track student's learning can use this information to automatically detect groups of students with similar learning patterns.

By clustering students based on keyword customization, it might be possible to determine if there are groups of students that frequently customize similar subjects. Such a result could be very useful in searching, cataloging, and discovering the huge number of students that are currently online.

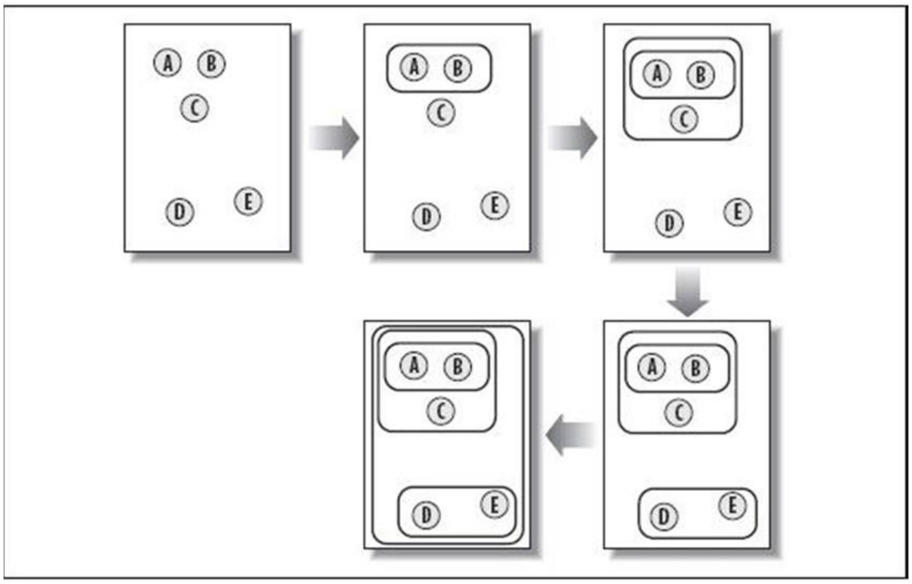

Fig. 1. Hierarchical clustering in action

Hierarchical clustering builds up a hierarchy of groups by continuously merging the two most similar groups. Each of these groups starts as a single item, in this case an individual student. In each iteration this method calculates the distances between every pair of groups, and the closest ones are merged together to form a new group. This is repeated until there is only one group.

In the Fig. 1, the similarity of the items is represented by their relative locations, the closer two items are, the more similar they are. At first, the groups are just individual items. In the second step, you can see that A and B, the two items closest together, have merged to form a new group. In the third step, this new group is merged with C. Since D and E are now the two closest items, they form a new group. The final step unifies the two remaining groups.

To find users who are similar to one user, cluster models divide the users' base into many segments. Using a similarity metric, a clustering algorithm groups the most similar users together to form clusters or segments. These algorithms typically start with an initial set of segments, which often contain one randomly selected customer each. They then repeatedly match students to the existing segments. The algorithm's goal is to assign the user to the segment containing the most similar users. And student in the one group has the same keyword customization.

5. Recommendation based on clustering

For users, sometimes there is the problem of poor accuracy of the information, and the model could not accurately grasp the user's real needs only through keywords, so the information push by the model and the real needs of users exists differences.

And in most cases, users do not know which words should be customized, so we should extend the function of recommending keywords to the users automatically. Therefore, in order to more accurately express the user's information needs, the system requires a recommend function, by exploring the user's behaviors and other users with similar behaviors. To solve these issues, you need to score the keywords by producing a score that rating.

Most recommendation algorithms start by finding a set of users whose action and rated items overlap the user's rated items [3,4,5]. The algorithm aggregates items from these similar users, eliminates items the user has already rated, and recommends the remaining items to the user.

The task of the recommendation algorithm concerns the prediction of the target user's rating for the target item that the user has not given the rating, based on the user's ratings on observed items as shown in table 1.

Table1. Users and keywords rating score

UID	K1	K2
User1	*5*	*4*
User2	*4*	*5*
User3	*0*	*N/A*

What will the user3 rate to the keyword2? We want to determine how much better one keyword is liked than another. One way to measure this differential is simply to subtract the average rating of the two keywords. In turn, this difference can be used to predict another user's rating of the keywords by given their rating of the other.

Consider three users and two keywords, user1 gave keyword1 a rating score of 5 while user1 gave keyword2 a rating score of 4. We observe that keyword1 is rated more than keyword2 by 5−4=1 points, and user2 rated keyword2 more than keyword1 by 5-4=1, thus the average value between keyword1 and keyword2 is ((5-4) + (4-5))/2=0, we could predict that user3 will give keyword2 a rating of 0+0=0. This is the slope one algorithm [1, 2], it is simple and efficient.

There are N individuals score with keywords A and B, R(A—>B) indicates that the points average difference by these N individuals. And M individuals score with keywords B and C, R(B—>C) indicates that the points average difference by these M individuals. And now user1 give the score of the keywordA is RA, give the keywordC is RC, then user1 give the score to keywordB might be.

In order to enhance the computing speed and scoring accuracy, we recommend based on clustering. We cluster users first, and each user in one cluster customized keywords are consistent. Inside the scoring system, users for each keyword has its own score, so we can count the scores for each user, which can be defined as the weighted average score of each cluster's rating score for one keyword. Our scoring system carry out the recommendation of the relevant knowledge based on the scores of the cluster. The recommendation here recommended knowledge to all users inside the cluster. The table will be amended as follows:

Table2. Clusters and keywords rating score

GID	K1	K2
Group1	5	4
Group 2	4	5
Group 3	0	0

6. Experiment method

The effectiveness of the method can be measured precisely. To do so, we test our schemes over the data set by knowledge management system. We used enough evaluations to have a total of 5,000 users as a data set. The data is collected from personalize customization system where ratings range from 0.0 to 5.0 for each keyword.

Table 3. Customize table

UID	K1	K2	K3	K4	K5	K6
User1	1	1	1	0	0	1
User 2	1	1	0	0	0	1
User 3	1	1	1	1	0	0
User 4	0	0	0	1	1	0
User 5	0	0	0	0	0	0

The system constructs the customize table, as shown in table3, of which the number "1" indicates that users have customized the corresponding keywords, the number "0" indicates that users does not customize the corresponding keywords. We should use cluster algorithm to analyze the customize keywords before rating scores calculation.

After clustering, we can obtain the following relationship between the user groups and his customization words, as shown in table 4.

Table 4. Users and new customize keywords

GID	List of keywords	UID
Group1	K1,K2,K3,K4, K6	User1, User2, User3
Group 2	K1,K2, K3,K4,K6	U4
Group 3	K2,K3,K4	U5

Based on the score algorithm, we could get the users rating scores table in table 5.

Table 5. Groups and keywords rating score

UID	K1	K2	K3	K4
Group1	5	5	3	N/A
Group2	3	5	0	5
Group3	5	0	N/A	5

For example, we have three groups and four keywords, among them, N/A indicates that the group user did not customize this keyword actively, but that does not mean the group would not need this keyword, maybe they did not know that they want to customize the keyword. Then we should calculate the score difference between two keywords, we should get the following matrix as shown in table 6.

Table 6. Scores and frequency

	K1	K2	K3	K4
K1	N/A	3/3	5/2	1/2
K2	-3/3	N/A	7/2	-5/2
K3	-5/2	-7/2	N/A	-5/1
K4	-2/2	5/2	5/1	N/A

Among the table, such as "1/2", 1 for the score difference between two keywords and 2 represents the number of groups who have given the score. Consider the weighted slope one algorithm, we should record how many people have rated the scores. This combined with the first step will get the group ratings matrix.

We should get the prediction of the target group's rating for the target keywords. When the calculation rating score is equal to or greater than 3, this means that the group will choose to customize the keywords directly. The system could push the content concern with this keyword to the user in the group automatically. When the score is greater than 0 and less than 3, this means that the system could not determine whether the user will customize the keyword or not, so this keyword should be recommended to the group, allow user in the group choose to score it. When the score less than 0, this means that the group will have no interest in the keywords, the system may not recommend it to the group.

7. Conclusion

This paper introduces collective intelligence on energy economic knowledge management system, and proposes a new dynamic programming approach of energy economic keyword recommendation by analyzing the personal behaviors. Slope one algorithm join with clustering algorithm have been proved that they can greatly reduce the complexity of computational complexity.

Collective intelligence is based on crowd behavior research, and does the personalized recommendations for each user; so that users can accurately information they need more quickly. Energy economic knowledge management system must base on user's preferences and provide precise recommendations on knowledge learning. Recommendation must be real-time calculation, so as to allow the user access to the recommended content and give the feedback on the recommended contents in time.

References

[1] Daniel L, Anna M, Slope One Predictors for Online Rating-Based Collaborative Filtering, Proceedings of The Fifth SIAM International Conference On Data Mining 2005; 1:471-475.

[2] Daniel L, Scale and Translation Invariant Collaborative Filtering Systems, Information Retrieval 2005; 8:129-150.

[3] Herlocker J, Konstan J, Terveen L, Evaluating Collaborative Filtering Recommendation System, ACM Trans. On Information System 2004; 22:5-53.

[4] Kin B, Li Q, Park C, Et Al., A New Approach for Combining Content-Based and Collaborative Filters, Journal of Intelligent Information Systems 2006;27:79-91.

[5] Slobodan Vucetic, Zoran Obradovic, Collaborative Filtering Using a Regression-Based Approach, Knowl. Inf. Syst 2005; 7: 1-22.

[6] Demirbas. Energy Issues in Energy Education. Energy Education Science and Technology Part A 2011; 27: 209-220.

[7] Bell P, Scientific arguments as learning artifacts: Designing for learning from the web with KIE. International Journal of Science Education, 2000, 22(8), 797-817.

[8] Joo Y J, Bong M, Choi H J, Self-efficacy for self-regulated learning, academic self-efficacy, and Internet self-efficacy in Web-based instruction. Educational Technology Research and Development, 2000, 48(2), 5-17.

[9] Moore M G, Kearsley G, Distance education: A systems view of online learning. Cengage Learning. 2011

[10] Aydin S, A review of research on Facebook as an educational environment. Educational technology research and development, 2012, 60(6), 1093-1106.