www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 5 May 2017, Page No. 21405-21409
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i5.36

Novel Weighted Hybrid Approach in Recommendation Method

Mr. Avadhut D. Wagavkar, Mrs. S.S.Vairagar
Pune University, Siddhant College of Engineering,
Sudumbare, Talegaon, 412109 Pune India
adwagavkar712@gmail.com
Pune University, Siddhant College of Engineering,
Sudumbare, Talegaon, 412109 Pune India
Vairagarss@gmail.com

Abstract: ***In modern E-Commerce it is not easy for the customers to find the best goods of their interest as there are millions of products available online. Recommendation systems are one of information filtering systems forecasting the items that may be additional interest for user within a big set of items on the basis of user's interests. This System utilizes the Collaborative filtering, which offers a few recommendations to users on the basis of matches in behavioral and useful examples of users and furthermore demonstrates comparable affection and behavioral examples with those users. The paper presents an approach for Recommendation System to generate meaningful recommendations to a collection of users for items or products that might interest them. This approach uses weighted hybrid recommendation system which combines content based recommendation system and knowledge based recommendation system in order to increase the overall performance of the system. The main idea is using multiple recommendation techniques to suppress the drawbacks of the traditional techniques or an individual technique in a combined model. The paper presents a system to improve the accuracy of recommendation in big data application.***

Keywords: Recommender System, Big Data, Clustering, Collaborative Filtering, Preferences, Weighted Hybrid Recommendation System.

1. Introduction

The development of the Internet provided more ways for people to interact but also a place here they could find information about almost everything and anything. Recommendation systems can be considered a way of combining these two aspects in order to help people find the information they need or something they would be interested in. Recommendation systems are used in various online applications from e-Commerce to search engines. There are a number of techniques used to implement recommendation systems, each with its advantages and disadvantages. Hybrid systems used to combine two or more of these techniques in order to obtain accurate results. Recommendation systems development was driven by e-Commerce but there are also other applications for them such as search results and news portals customization. Hybrid techniques were implemented to overcome some of the deficiencies in the traditional techniques. The deficiencies include performance aspects, but also trust security and privacy issues. The most commonly used technique in recommendation systems is collaborative filtering [4].

The other techniques that are used to classify recommendation systems are presented here. The classification is based on data that is collected automatically (background) and the data that is introduced by the users (input). Collaborative filtering techniques are based on the ratings that users gave to the products. These ratings are used to find similar users and based on that community of users, products are recommended. Content-based methods are utilized to filter information in view of a user profile. The user profile is worked by discovering habits of the users in the data available. Utility-based and knowledge-based systems are comparable as they are not in light of past exchanges but rather on user needs. Other factors apart from item characteristics are taken into account. Both utility-based and knowledge-based systems make a user profile that mirrors the requirements of the user. Amongst the metrics that asses the accuracy of a recommendation system, identify precision and recall. Precision is defined as the ratio of items predicted correctly to the total number of items predicted. The ratio of items predicted correctly to the total number of items that can be selected called Recall. In other words, precision is defined as the probability that a recommended item is relevant while recall is the probability that a relevant item is recommended.[3]

This technique is a combination of content based recommendation algorithm and knowledge based algorithm. It enhances the performance of the system. Hybrid recommender systems combine multiple recommendation techniques in order to increase the overall performance. The main idea is using multiple recommendation methods to clear the drawbacks of an individual technique in a combined model [6].

A weighted hybridization strategy combines the recommendations of multiple recommendation systems by computing weighted sums of their scores. These score are hybridized by using a uniform weighting scheme.

2. Content Based Recommendation

A content based recommendation technique works on the user preferences that is likes and dislikes given by any user to items or products and the user profile. Here it will only consider likes given by user. The intuition of a content-based approach is that two objects having similar contents are similar. Generally, it uses keywords to describe the content of objects. The keyword sets of an object can be extracted from the content by several approaches. The most used one is the term frequency-inversed document frequency (TF-IDF). Term frequency is the appearing frequency of a word in a document. Intuitively, the more times of a word appear; the more important the word is to the document. However, some words appear in too many documents, so these words cannot describe the topic of document well. For instance, the words “smartphone” and “iPhone” appears in a same document. Both words are considered important to describe the topic of the document. However, the word smartphone appears in much more documents than the word iPhone does. Thus, we consider that iPhone can describe the topic of the document better than smartphone does. Therefore inversed document frequency (IDF) is introduced to represent the importance of words, which represents the frequency that a word appears in all documents. The TF-IDF value of a word to a document is computed as follows:

$$\text{Tf}\ (w,\ d) = \frac{n_{w,d}}{n_d}$$

$$\text{Idf}\ (w,\ d) = \log \frac{N}{N_w}$$

$$\text{Tf-idf}\ (w,\ d) = \text{tf}\ (w,\ d) \times \text{idf}\ (w,\ d)$$

Where nw,d is the number of times a word w appears in a document d, nd is the total number of the document’s words, N is the total number of documents in the corpus, and Nw is the number of documents in which w appears. The keywords having high TF-IDF value are considered important to the document. Note that a keyword’s TF-IDF value is not same in different documents. A word that is an important keyword in one document may be not important in another.

After calculating the top k keywords’ TF-IDF values, these k keywords constitute keyword sets of the object. We can compare the similarity between two objects’ keyword sets to calculate the similarity between their contents. Generally speaking, the more same keywords two objects have, the more similar they are. Moreover, the appearance count of keywords is also an important aspect. A frequently-used method to calculate two keyword sets’ similarity is cosine similarity. The cosine similarity between two objects is given as,

$$\text{Sim (a, b)} = \frac{\sum_{i=1}^{n} ai \times bi}{\sqrt{\sum_{i=1}^{n} a_i^2} \times \sqrt{\sum_{i=1}^{n} b_i^2}}$$

Where n is the number of keywords, ai and bi are the ith keyword’s, a, b are weight of objects respectively. The weight is usually defined as a keyword’s TF-IDF value.

3. Knowledge Based Recommendation

A knowledge based recommendation technique works on the set of requirements of user and the product description. Products feature and category are compared with the user’s interest and category respectively.
Steps of algorithm are below,

3.1 Capture user Knowledge by a Knowledge based approach:

An active user can give his/her requirement about candidate services.

3.2 Calculate personalized ratings and generates recommendations:

The personalized ratings of each candidate service for the active user can be calculated by using multi attribute utility theory (MAUT), which evaluates each item with regard to its utility for the user. Each item is evaluated according to a predefined set of dimensions that provide an aggregated view on the basic item properties. The user-specific item utility is calculated on the basis of following formula.

$$\text{Utility (P)} = \sum_{j=1}^{dimensions} interest(j) \times Contribution(p,j)$$

Where, index j iterates over the number of predefined dimensions, interest (j) denotes a user’s interest in dimension j, and contribution (p, j) denotes the contribution of item p to the interest dimension j.

4. Weighted Hybrid Recommendation

This technique is a combination of content based recommendation algorithm and knowledge based algorithm. It enhances the performance of the system. Hybrid recommender systems combine two or more recommendation techniques in order to increase the overall performance. The main idea is using multiple recommendation methods to clear the drawbacks of an individual technique in a combined model.

A weighted hybridization strategy combines the recommendations of two or more recommendation systems by computing weighted sums of their scores. These score are hybridized by using a uniform weighting scheme. Thus, given n different recommendation functions reck with associated relative weights βk.

$$Rec_{weighted}(u,i) = \sum_{k=1}^{n} \beta_k \times Rec_k(u,i)$$

5. Agglomerative Hierarchical Clustering (AHC) Algorithm

Assume there are n services. Each service is initialized to be a cluster of its own. At each reduction step, the two most similar clusters are merged until only K (K < n) clusters remains.

Input:

A set of services S= {S1… Sn},

A characteristic similarity matrix D= [di, j] nxn the number of required clusters k.

Output:

Dendrogram for k=1 to S.

1. Ci = {Si}, ∀ i;

2. dCi,Cj = di,j , ∀ i, j;

3. For k = S downto K

4. Dendogram k = {C1…, Ck};

5. Lm = dCi, Cj;

6: Cl = Join (Cl, Cm);

7. for each Ch∈S

8. If Ch ≠ Cland Ch ≠ Cm

9. dCl,Ch = Average(dCl,Ch , dCm,Ck);

10. Endif

11. Endfor

12. S = S – {Cm};

13. Endfor

6. Literature Review

"Service-generated Big Data and Big Data-as-a-Service: An Overview" [1] with the prevalence of service computing and cloud computing, more and more services are emerging on the Internet, generating huge volume of data, such as trace logs, QoS information, service relationship, etc. The overwhelming service-generated data turn out to be too vast and complex to be successfully handled by conventional methodologies. How to store, manage, and create values from the service-oriented big data become an important research problem. Then again, with the increasingly extensive measure of information, a single infrastructure which gives normal usefulness to overseeing and analyzing different types of service-generated big data is urgently required. To address this challenge, this paper provides an overview of service-generated big data and Big Data-as-a-Service. Initial, three sorts of service-generated big data are misused to improve system execution. Then, Big Data-as-a-Service, including Big Data Infrastructure-as-a-Service, Big Data Platform-as-a-Service, and Big Data Analytics Software as- a-Service, is employed to provide common big data related services (e.g., getting to benefit produced big data and data investigation results) to users to improve proficiency and reduce cost.

Zibin Zheng, Hao Ma, Michael R. Lyu, and Irwin King "QoS-Aware Web Service Recommendation by Collaborative Filtering" [2] with increasing presence and adoption of Web services on the World Wide Web, Quality-of-Service (QoS) is becoming important for describing nonfunctional characteristics of Web services. This paper presents a collaborative filtering approach for predicting QoS values of Web services and making Web service recommendation by taking advantages of past usage experiences of service users.

This paper proposes a user-collaborative mechanism for past Web service QoS information collection from different service users. Then, based on the collected QoS data, a collaborative filtering approach is designed to predict Web service QoS values. Finally, a prototype called WSRec is implemented by Java language and deployed to the Internet for conducting real-world experiments. To study the QoS value prediction accuracy of our approach, 1.5 million Web service invocation results are collected from 150 service users in 24 countries on 100 real-world Web services in 22 countries. The experimental results show that our algorithm achieves better prediction accuracy than other approaches. Our Web service QoS data set is publicly released for future research.

"An efficient hybrid algorithm based on modified imperialist competitive algorithm and K-means for data clustering" [3] clustering techniques have received attention in many fields of study such as engineering, medicine, biology and data mining. The aim of clustering is to collect data points. The K-means algorithm is common techniques used for clustering. However, there results of K-means depend on the initial state and converge to local optima. Keeping in mind the end goal to conquer nearby optima obstacles, a considerable measure of studies have been done in bunching. This paper presents an efficient hybrid evolutionary optimization algorithm based on combining Modify Imperialist Competitive Algorithm (MICA) and K-means (K), which is called K-MICA, for optimum clustering N objects into K clusters. Then new Hybrid KICA algorithm is tested on several data sets and its performance is compared with those of MICA, ACO, PSO, Simulated Annealing (SA), Genetic Algorithm (GA), Tab Search (TS), Honey Bee Mating Optimization (HBMO) and K-means. The recreation comes about demonstrate that the proposed developmental enhancement algorithm is vigorous and reasonable for taking care of data clustering.

"Using multidimensional clustering based collaborative filtering approach improving recommendation diversity."[4] Li et al. proposed to incorporate multidimensional clustering into a collaborative filtering recommendation model. Background data in the form of user and item profiles was collected and clustered using the proposed algorithm in the first stage. Then the poor clusters with similar features were deleted while the appropriate clusters were further selected based on cluster pruning. At the third stage, an item prediction was made by performing a weighted average of deviations from the neighbor's mean. Such an approach was likely to trade-off on increasing the diversity of recommendations while maintaining the accuracy of recommendations.

7. Proposed System

Proposed weighted hybrid system provides recommendations to user by using clustering based hybrid recommendation system for semantic clusters. First of all the user interact with data processing unit. Data is collected from users in the form of big data; data will be in the form of users purchase information, rating information, according to user's area of interest, product key features, purchase history. After the data collection data processing step takes place where semantic clustering techniques are applied to collected data. Complete data is divided into the number of clusters. For this purpose Agglomerative Hierarchical Cluster (AHC) algorithm is used. These clusters are input to the collaborative filtering techniques. Weighted Hybrid recommendation system is used to provide final recommendations to the users.

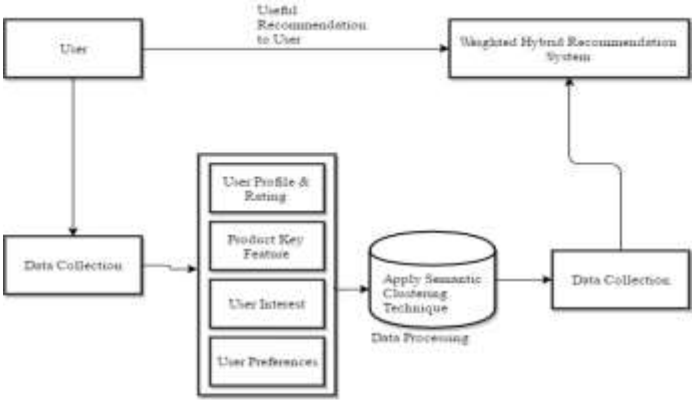

Figure 1: Architectural Design

8. Mathematical Model

Let WHRS be a weighted hybrid Recommendation System.

WHRS= {I, S, Sclust, FCBPred, FKBPred, Pd, Uf, Pr, Uin, Up, Robj, O| Φp}

Where,

I =Input – big data-Service Clusters

S=User session

S= {S1, S2... Sn}

Sclust - List of semantic clusters

List< Sclust >= AHC (Bigdata)

FCBPred = Content based recommendations on Uf and Up

FCBPred = Sclust (Uf + Up)

Uf - User profile

Up - User preferences

FKBPred = Knowledge based recommendations on Uf, Uin and Pd.

FKBPred = Sclust (Uf + Uin + Pd)

Uf - User profile

Uin - User's attributes and area of interest.

Pd - Product with its key features or description

Pr - Product rating's given by individual users

Up - User preferences

Robj - Recommended product

O= Output = R.

R- Weighted Hybrid Recommended List

R1 = FCBPred

R2 = FKBPred

R =Σ (R1 + R2)

Return R.

9. Result Analysis

In order to calculate accuracy Mean Absolute Error (MAE) is calculated as shown in the following equation.

$$MAE = \frac{\sum_{i=1}^{n} |r_{a,t} - P(u_a, s_t)|}{n}$$

Where, n is the total no. of items or products or services. In case of item based collaborative filtering, r (a, t) is the actual rating given by user to the product. P (ua, st) is the predicted ratings. In case of hybrid recommendation system, r (a, t) is the total no of items who has been rated as well as preferred by the user and P (ua, st) is the predicted ratings. Low MAE values represent high accuracy. For the simplicity predicted values are calculated as follows:

$$P_{u_a,s_t} = \overline{r}_{s_t} + \frac{\sum_{sj \in N(s_t)} \left(r_{u_a,sj} - \overline{r}_{sj}\right) \times R\_sim'(s_t, s_j)}{\sum_{sj \in N(s_t)} R\_sim'(s_t, s_j)}$$

Calculated MAE values are represented in the table. From the table it is clear that the proposed system is having low mean absolute error it means the proposed system i.e. weighted hybrid recommendation system is more accurate as compared to existing system.

Table 1: Comparison Table

Cluster Size	MAE (User Based)	MAE (Weighted Hybrid)
C1 (37)	1.189	0.216
C2 (46)	0.956	0.195
C3 (80)	0.825	0.225
C4 (63)	0.619	0.206
C5 (146)	0.993	0.198
C6 (292)	0.969	0.202

Graphical representation of the result is as shown in the following figure. Red bars represent User based system and green bars represent the proposed system. X-axis represents the cluster size n Y-axis represents the accuracy.

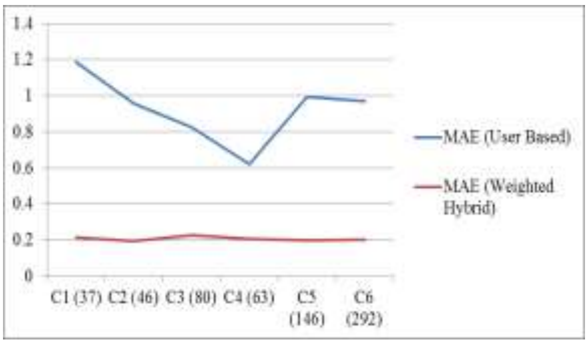

Figure 2: Comparative Result

Figure 1: Testing data- load current (amperes)

References

[1] W. Dou, X. Zhang, J. Liu, and J. Chen, "Hire Some-II: Towards Privacy-Aware Cross-Cloud Service Composition for Big Data Applications," IEEE Trans. Parallel and Distributed Systems, 2013.R. Caves,

Multinational Enterprise and Economic Analysis, Cambridge University Press, Cambridge, 1982. (book style)

[2] Akshita, Smita "Recommender System: Review" International Journal of Computer Applications (0975 -8887) Volume 71– No.24, June 2013 38H.H. Crokell, "Specialization and International Competitiveness," in Managing the Multinational Subsidiary, H. Etemad and L. S, Sulude (eds.), Croom-Helm, London, 1986. (book chapter style)

[3] Shanshan Cao "A Hybrid Collaborative Filtering Recommendation Algorithm for Web-based Learning Systems" 2015 International Conference on Behavioral, Economic, and Socio-Cultural Computing.

[4] Khushboo R. Shrote, Prof. A.V.Deorankar "Review Based Service Recommendation for Big Data" International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB16)

[5] Iman Barjasteh, Rana Forsati, Dennis Ross, Abdol-Hossein Esfahanian, Hayder Radha, Fellow, and IEEE "Cold-Start Recommendation with Provable Guarantees: A Decoupled Approach"

[6] R. Burke, "Hybrid Recommender Systems: Survey and Experiments," User Modeling and User-Adapted Interaction, vol. 12, no. 4, pp. 331-370, 2002

[7] G. Salton, "Automatic Text Processing". Addison-Wesley, 1989. A. Chu, R. Kalaba, and K. Spingarn, "A Comparison of two Methods for Determining the Weights of Belonging to Fuzzy Sets," Optimization Theory and Applications, Springer vol. 27, no. 4, pp. 531-538,1979.

[8] Profiles Akihito NAKAGAWA and Takayuki IT0 "An Implementation of a Knowledge Recommendation System based on Similarity among Users"

[9] T.N. Chiranjeevi and R.H. Vishwanath21 Department of Computer Science and Engineering, Sambhram Institute of Technology, India "Prs: Personnel Recommendation System For Huge Data Analysis Using Porter Stemmer" ISSN: 2229-6956 (Online) ICTACT Journal on Soft Computing, April 2016, Volume: 06, Issue: 03 1235

[10] Dept. of CA, Vasavi College of Engineering, Hyderabad-31, India Dr. T.Adi Lakshmi, Dept. Of CSE, Vasavi College of Engineering, Hyderabad-31, India "Recommendation Systems: Issues and challenges" Soanpet .Sree Lakshmi et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5771-5772 Soanpet .Sree Lakshmi.

Author Profile

1] Avadhut Wagavkar received the B.E. (Computer Sci & engg.) from Adarsh Institute of technology and research center, Vita, Sangli, in 2013 and Pursuing M.E. (Computer Engg.) degree from Siddhant College of engineering Sudumbare, Talegaon, Pune.

2] Mrs. S.S. Vairagar Working as a Head of Department of computer engineering at Siddhant College of engineering Sudumbare, Talegaon, Pune. She has 11th years of experience in teaching.