International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 02 | Feb -2017 www.irjet.net p-ISSN: 2395-0072

Providing highly accurate service recommendation for semantic clustering over big data

Neha D. Patil[1], Dr. D. S. Bhosale[2]

[1]PG Student, Ashokrao Mane group of institution, Vathar

[2]Associate Professor, Ashokrao Mane group of institution, Vathar

---

Abstract - *Numerous approaches have been proposed to provide recommendations. Manifestly, recommendation system has a variety of properties that may need experiences of a user, such as user prediction, rating, trust, etc. On the internet, where the number of choices is enormous, there is a need to filter, prioritize and efficiently deliver relevant information to mitigate the problem of many internet users. Recommended systems are one of information filtering systems, estimating the items that may be of additional interest to user within a big set of items based on a user's interests. Recommended systems are currently useful in both the research and in the commercial areas. The paper presents an approach for Recommended System to generate meaningful recommendations of a collection of users for items that might interest them. This approach uses adaptive recommender system which combines two recommendation techniques to increase the overall performance. The main aim of using multiple recommendation techniques to overcome the drawbacks of the traditional techniques in a combined model. The anatomy is based on the hierarchy and input/output relations of recommenders. The present system improves the speed and accuracy of recommendation in big data application.*

***Key Words*: *Adaptive Recommendation System*, clustering, data mining and Big data.**

1.INTRODUCTION

Big Data relates large-volume, growing and complex data sets with multiple and independent sources. In Big Data applications, data collection has increased terribly and it is beyond the ability of commonly used software to capture, manage, and process that data [3]. The most crucial challenge to Big Data applications is to inspect the large volumes of data and get useful information or knowledge for future actions. Service users nowadays encounter unrivalled difficulties in finding ideal services from the enormous services. These days, it is common for people to choose web as the platform to buy or sell something. Therefore, there exist many online shops in different forms, varying from private websites to eCommerce forums. This leads to both advantages and disadvantages for customers in different ways [1] The main advantage is that a customer has more options to buy from. At the same time, it can also have the drawback, because with many options customer will face difficulty to choose one single product keeping in view various criteria e.g. which shop has good customer service, and who offers the best price. Therefore, the big issue is that there is no one-stop place to search wide information about e-Commerce. The information which is required related to online selling and buying includes list of products, list of online shops and a set of recommendations about choosing product and shop.

Recommender system is information filtering system that deals with the problem of information excess [7]by filtering vital information out of large amount of dynamically generated information as per user's preferences, observed behavior about item or interest [9]. Recommender system has capacity to forecast whether a user would select an item on the user's profile. Collaborative filtering (CF) techniques such as item-, user- and utility-based are the governing techniques applied in RSs. However, traditional CF techniques are sound and have been successfully applied in many RSs. They face two main challenges in big data application:1) to explore useful recommendations from so many services and 2) to take a decision within limited time. A critical step in traditional CF algorithms is to compute likeness between every pair of users and/or services which may take long a time, also beyond the processing capability of current RSs. The ratings of dissimilar users or services may influence the accuracy of predicted ratings. One solution is to reduce the number of services that need to be processed in real time. Clustering are such techniques that can decrease the data size by a large factor by grouping similar services together. Therefore, the paper proposes a clustering and collaborative filtering with adaptive recommendation technique. Clustering is approach that separate big data into manageable partitions [4]. Besides, since the ratings of similar services within a cluster are more pertinent than that of dissimilar services, the recommendation accuracy based on users' ratings may be enhanced. Despite the success of filtering techniques, they exhibit cold-start, sparsity and scalability problem. This paper proposes an adaptive recommendation system that combines item- and knowledge-based filtering techniques to increase the accuracy and performance of RSs.

2.RELATED WORK

Li et al [6] proposed to integrate multidimensional clustering into a collaborative filtering recommendation model. Background data in the form of user and item profiles were collected and clustered using the proposed algorithm in the first stage. Then the poor clusters with similar features were deleted while the suitable clusters were selected based on cluster pruning. At the last stage, an item prediction was done by performing a weighted average of the deviations from the neighbor's mean. An approach like this was likely to trade-off on increasing the variety of recommendations while maintaining the accuracy of the recommendations.

Pazzani et al. [12] developed an intelligent agent that tries to evaluate which pages in the website will appeal a user by applying a naïve Bayesian classifier. The agent allows a user to give training instances by rating different pages as either hot or cold. Jennings and Higuchi describe a neural network that models the interests of a user in a Usenet news environment.

Pham et al. [8] proposed to use network clustering technique for social network of users to recognize their neighborhood. And then use the conventional CF algorithms to explore the recommendations. This work relies on social interaction between users.

Simon et al. [10] used a high-dimensional parameter-free, divisive hierarchical clustering algorithm that needs only implicit feedback on past user purchases to find the relationships within the users. Depending on the results of clustering, products of high interest were suggested to the users. But, suggested feedback does not always provide accurate information about the user's choice.

In Ziegler, et al. [5], a hybrid collaborative filtering approach was proposed to capitalize on bulk taxonomic information, designed for extracting product classification to address the data sparsity problem of CF recommendation, based on the generation of profiles via inference of super-topic score and topic diversification.

A hybrid recommendation technique is also proposed in Ghazantar and Pragel-Benett [4] and this uses the content-based profile of the individual user to find similar users which are used to make predictions.

Clustering techniques have received attention in many fields of study such as engineering, medicine, biology and data mining. The goal of clustering is to collect data points. The K-means algorithm is one of the most common techniques used for clustering. But, their results of K-means depend on the initial state and converge to local optima. To overcome local optima obstacles, a lot of studies have been done in clustering. T. Nickname, Taherian Fared proposed [11] a paper that 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 optimizing clustering N objects into K clusters. Then new Hybrid K-ICA 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 simulation results show that the proposed evolutionary optimization algorithm is robust and suitable for handling data clustering.

With the fact of service computing and cloud computing, more and more services are appearing on the Internet, generating huge volumes of data, such as trace logs, QoS information, service and relationship etc. The enormous service-generated data has become too large and complex to be effectively processed by conventional approaches. How to store, manage, and create value from the service-oriented big data has become an important research problem. On the other hand, with the increasingly large amount of data, a single infrastructure which provides common functionality for managing and analyzing different types of service-generated big data is urgently required.

To address this challenge, Zibin Zheng, Jieming Zhu and Michael R. Lyu [2] proposed a paper that provides an overview of service-generated big data and Big Data-as-a-Service. First, three types of service-generated big data are exploited to enhance system performance. 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. accessing the service generated big data and data analytics results) to users to enhance efficiency and reduce cost.

3.PROPOSED WORK

Here, designed system uses an adaptive recommendation system which is making automatic predictions about the interests of a user by gathering preferences from many users. The aim of this system is to recommend new items to the user or forecast the utility of a certain item, based on user's previous likings and on the opinions of other like-minded users. Adaptive Recommendation system dominates content based recommender system and collaborative filtering recommender, as the relative accuracy of the recommender is comparatively high.

A. System Architecture

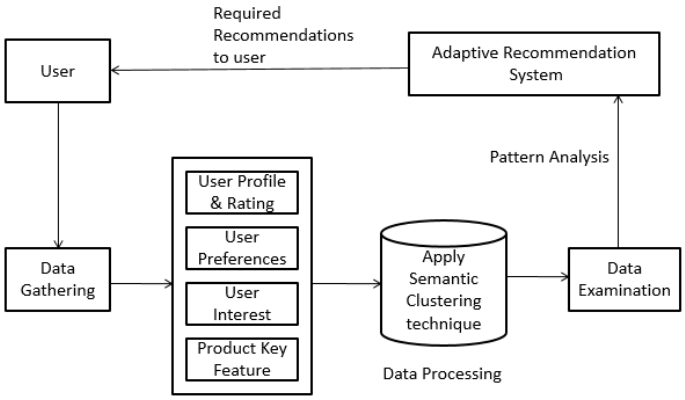

Fig 1: Architecture of Adaptive recommendation system for semantic cluster in big data

The proposed architecture consists of two modules.

1. Admin Module

Admin can login and manage the categories of products. The admin user can add/update product information, images and description. Also, the admin has right to approve the order requested by visitor, generate invoices and pass to dispatch team. Admin module has the option to add/update advertisements. Also, admin can provide the product details, advertise details and order details to visitor through web services.

Cluster Generation (Compute Description Similarity and Functionality Similarity):

A. Description similarity: Description similarity is computed by a Jaccard similarity coefficient (JSC) which is a statistical measure of similarity between sample sets.
B. Functionality similarity: Functionality similarity is also computed by a Jaccard similarity coefficient (JSC), similarly as description similarity.

2. Visitor Module

In the proposed system visitor module is the second module. First, visitor can register and login to the web portal. Visitor user can view product list by categories. Also, the visitor module has the option to view details of product and purchase the product and notify by mail. Visitor can give rating to the product and he can give a prediction about that product.

Service Recommendation using (Adaptive Collaborative Filtering Techniques)

The recommendation system builds a database (user-item matrix) of preferences for items by user. It then matches users with preferences by calculating similarities between their profiles to make recommendations. Such users build a group called a neighborhood. A user gets recommendations to those items that he has not rated before, but that was already positively rated by users in his neighborhood. Recommendation that is produced can be either prediction or recommendation. Prediction is a numerical value, Rij, expressing the predicted score of item j for the user i, while recommendation is a list of top N items that the user will like the most. The adaptive system combines content-based filtering and Item-based collaborative filtering.

In content-based filtering technique, recommendation is made based on the user profiles using features extracted from the content of the items evaluated in the past. Items that are mostly related to the positively rated items are recommended to the user.

Item-based approach computes for each user-item correlation with all other items and aggregates for each user the ratings for item that are already highly correlated.

4.SCOPE OF THE WORK

The main goal of the proposed work is providing accurate recommendations to users.

1.To present the new adaptive algorithm to improve the Scalability, Accuracy, Memory consumption.

2.Present an approach that provides the recommendation to users even they are new in system by removing the cold-start problem from existing algorithm.

3.Present a system to improve the speed and accuracy of recommendation system in big data application.

5. CONCLUSION

In this paper, Recommendation techniques were studied and a new system is proposed, that uses an adaptive recommendation approach which combines content-based filtering and Item-based collaborative filtering. Adaptive algorithm improves the Scalability, Accuracy and Memory consumption. Also, removes cold-start, data sparsity and scalability problem. The new proposed system will perform clustering and provide accurate recommendations using adaptive approach.

6.REFERENCES

[1] NanangHusin," Internet user Behavior Analysis in online shopping on Indonesia", processing of the 2011 international conference on Advanced Computer Science and information System (ICASIS).

[2] Z. Zheng, H. Ma, M. R. Lyu, and I. King, "QoS-aware web service recommendation by collaborative filtering," IEEE Trans. Services Compute, Vol. 4, no. 2, pp. 140-152, Feb. 2011.

[3] M. A. Beyer and D. Laney, "The importance of 'big data: A definition "Gartner Inc., Stamford, CT, USA, Tech. Rep, 2012.

[4] T. C. Havens, J. C. Bezdek, C. Leckie, L. O. Hall, and M. Palaniswami, "Fuzzy c-means algorithms for very large data," vol. 20, no. 6, pp. 1130-1146, Dec. 2012.

[5] Ziegler CN, Lausen G, Schmidt-Thieme L. Taxonomy-driven computation of product recommendations in: Proceeding of the 13th international conference on information and knowledge management (CIKM 04), Washington DC, USA;2004

[6] X. Li and T. Murata, "Using multidimensional clustering based collaborative filtering approach improving recommendation diversity," in Proc. IEEE/WIC/ACM Int. Joint Conf. Web Intell. Intell. Agent Technol., Dec. 2012, pp. 169-174.

[7] Konstan JA. Riedl J. Recommender system: from algorithms to user experience. User model User-Adapt Interact 2012.

[8] M. C. Pham, Y. Cao, R. Klamma, and M. Jarke, "A clustering approach for collaborative filtering recommendation using social network analysis," J. Univ. Compute. Sci., vol. 17, no. 4, pp. 583-604, Apr. 2011.

[9] Pan C. Li W. Research paper recommendation with topic analysis. In computer design and Application, IEEE 2010;4. pp, v4-264.

[10] R. D. Simon, X. Tengke, and W. Shengrui, "Combining collaborative filtering and clustering for implicit recommender system "in Proc. IEEE 27th Int. Conf. Adv. Inf. Netw. Appl., Mar. 2013, pp. 748-755.

[11] T. Nickname, E. Taherian Fared, N. Pourjafarian, and A. Rousta, "An efficient algorithm based on modified imperialist competitive algorithm and K-means for data clustering," Eng. Appl. Artif. Intell., vol. 24, no. 2, pp. 306-317, Mar. 2011.

[12] Pazzani MJ A framework for collaborative content-based and demographic filtering. Artific Intell Rew 1999.