

IMPROVEMENT OF RECOMMENDER SYSTEMS CONSIDERING BIG DATA OF USERS' COMMENTS ON CHOSEN ITEMS

Kh. Esfandiari^{1,*}, A.R. Honarvar¹, Sh. Aghamirzadeh¹

¹Department of Electrical and Computer Engineering, Safashahr Branch, Islamic Azad University, Safashahr, Iran

Published online: 18 June 2016

ABSTRACT

Regarding to the increase in the online social networks services during the recent years, the recommender system has turned into an emerging research subject. Currently, regarding to the fast and consistent expansion of using the internet, the necessity of a recommender system for refining the large volume of data has increased greatly. The purpose of recommender systems is to provide a list of the interested items for the user and due to the increase in the current data volume, the previous used tools are nor suitable for processing this data volume; hence, having a system which can save and process the large data has turned into a problem. In this study, to solve the mentioned problems, a system is recommended using a model-based collaborative filtering refinement model which uses the Spark processing model in Hadoop context in order for more exact advising the user from the viewpoint of the users. The obtained results indicate that the current method will be more efficient and effective compared to the common recommendation methods.

Keywords: recommender systems; big data; sentiment analysis; hadoop; spark.

1. INTRODUCTION

The data volume in the world is increasing and its speed is more than the speed required to process it. Nowadays, humans have drowned in the informational range including: films,

Author Correspondence, e-mail: khadijaesfandiari@gmail.com

doi: <http://dx.doi.org/10.4314/jfas.v8i2s.141>



books, articles, news, products, etc. and it is difficult to identify what product is more suitable for them; then, a kind of technology is required which can process the available data and extract the most valuable items which are suitable and useful for the user. One of the most powerful technologies in this field are the recommender systems. Large volume of data on the internet can be the evangel of a context in which the user be able to simply find their desired and suitable items and if this available data is not properly managed, it will become a barrier to advance the purpose.

In this condition, the large volume of data and its increasing process make users face with the problem of finding appropriate item in appropriate time. The recommender systems are the smart systems which identify the preferences and interests of the users and recommend relevant items to them. This large volume of data allows us to use all available data optimally [1]. This data can be the score of users to the items, private information of users, their opinions about the items, content information of items, communications within the social networks and the geographic location of the user; in this study, in order to improve the recommender systems, we have used the users' textual reviews and call them {sentiment analysis}.

Over the years, extensive researches have been conducted with the aim of analyzing and classifying texts and data in which the purpose has been dedication of the predefined tags to the categories of documents based on the trained model. However, some newer researches have been done about the analysis of textual data to determine that how an individual represents his/ her "sentiments" about a certain subject. (For example, what are the feelings of the individual about that subject?) This has led to the development of sentiment analysis and the classification systems. Sentiment analysis and classification use the challenging techniques because the opinions can be stated in subtle and complex methods including using the informal language (slang), ambiguity, irony, humor and accent. Sentiment analysis and classification have been done for several reasons. For instance, to track the ups and downs of the way of dealing with a brand or product or compare the attitudes of online customers on a brand or product with another one, these are samples of a certain type of negative or positive comments on some issues. Also, it may be done in order to promote the management of

communication with customers and to help the other potential customers for conscious selection [2]. Due to the large volume of data, we use Spark processing engine to implement this recommender system; it focuses on the calculations within the memory; that is to say that in spite of the Ram capacity of the device, the calculations are done within the memory in Hadoop context which provides good possibility to store data in large scale by using cheap hardwares.

Apache spark is a cluster computing platform designed to be fast and general purpose. On the speed side, sparks extend the popular MapReduce model to efficiently support more type of computations, including interactive and stream processing. One of the main features spark offers for speed is the ability to run computations in memory, but the system is also more efficient than MapReduce for complex applications running on disk[3]. The aim of this study is to recommend the most similar item to the user regarding to the scores of the user to the items and also the users' comments on the items, so that by using the model-based collaborative filtering on the users' scores, it finds the most similar item to the item of the active user and in the next step, it tags the comments of users on the items using the survey method; and finally, the list obtained from the previous step is improved by eliminating the items which the users have had negative comments on them. Regarding to the increased data on web, and also the volume of the available data, the parallel plan and Spark processor in Hadoop context are used in order for fast processing this large volume of information.

This research includes the following hypotheses and after implementation and applying the system, we can provide an appropriate answer to these questions.

Is it possible to provide more exact advice by using the combination of algorithms of recommender system and survey?

Is it possible to satisfy the user by recommending the most similar item for him/ her?

2. Related Works

In the study [4] they provided music according to the priorities of the users for the recommender system. The methodology of the recommender system has been provided according to the sentiments metric. In this research, the "sentimeter-Br2" Dictionary has been used. The recommended framework is in the way that the user profile indicates the music

preferences of the user; and the music categories have been considered based on sadness, joy, anger, love and kindness; and the recommender system suggests music to each person according to the users' comments and his/ her own profile. User shares his/ her comments whenever he/ she logs on to the system; and these comments are analyzed using the defined dictionary and the intensity of sentiments is calculated and regarding to the negative or positive comments of the users, the system can select the type of recommended music to the user and in fact, the user's comments determine his/ her preferences and by using this recommender system and users' comments and their sentimental conditions, we can identify the way of changing of the people's sentiments in a period.

In [5] the problem of paper recommendation in Big Data scholarly was addressed. In this work an approximate approach for recommending papers to researchers based on local sensitive hashing proposed by converting the citations of papers to signatures and comparing these signatures against each other to detect similar papers according to their citations. A parallel and distributed aspect of the proposal is also discussed.

The study [6] has used a rank derivation approach and described the combination of collaborative filtering algorithm and the textual review of the users. The main idea of this approach is to extract the expressed preferences of the users which is a well-known problem for analyzing sentiments and the purpose is to implement the preferences on a number of rating scales which can be understandable by the current CF algorithms. The important duty in the context of rate inference is to determine the sentimental orientations (SO) and the strong points of the comment's words. For this reason, rate inference from a review, extracting the words of the comment in review and then the collection of SO of such words are for determining the dominant or average sentiments of the users.

In the research [7] a combination has been presented and the experimental work has been done with a new approach for the combined recommender system. The recommender systems are traditionally divided into three categories: content-based, collaborative filtering and combined method; the recommender systems have used a combination of both content-based and collaborative filtering methods to obtain better advices and benefit from their advantages. The research method is in the form that used an alternative formula for this experiment; so

that the content-based approach has been combined with a sentiment analysis to improve the results of recommendations. The final list of the advice only includes the cases with both similar meanings and also the user has liked them previously and have a positive tag in sentimental classifications.

In the study [8] a recommendation algorithm has been suggested based on the online reviews of users. This suggested algorithm is implemented in two steps of survey and recommender system and it says that unlike the traditional methods in which the recommendation to the user was deducted based on the similarity of the scores of users' preferences, in this recommender method, the difference between ranking and user's comments is analyzed to identify the user's preferences and a new different method is suggested in order to find out the users' preferences, the methods of prioritizing and recommending based on the users' comments and in this research, recommendations have been deducted based on the users' comments on each feature of items and preferences and according to the difference of the user's comments and the other users; in fact, a hidden idea has been provided in this method that if a lower score is presented about a feature, in order to eliminate the scoring habit of users, higher needs may be related to this feature.

3. Proposed system

Our suggested recommender system acts in the way that the recommender systems based on collaboration are used in combination with the sentimental analysis and implementing them on the tools of large data processing and using the users' comments for recommending them improves the results. So that in the first level, recommendation is done due to the collaboration-based algorithms and then, we improve the created list in previous step by analyzing the users' comments and the final list will be more exact and concentrated. The method used in this study uses the combination of two steps to advise the user. In the first stage, for each item and according to its ID all opinions of the users regarding to the dictionary-based approach and using the "sentiwordnet" are divided into two positive and negative classes. In the second stage, by means of the collaborative filtering algorithm (CF) and calculating the similarities among the users' and the active user items, the most similar items to the items of the active user are listed in a list to be recommended to the user. In this

stage, regarding to the comments obtained from the users of the first stage, the recommended list of the previous stage is combined with this and the items with the negative comments of the users are eliminated from the final recommendation list. Figure 1 shows the overview of the proposed method.



Fig.1.the architecture of the proposed system

The dataset used in this research was derived from Amazon Reviews obtained from “dataspan”² website with an 18 years period including 35 million review by March 2013. The part used in this study was obtained from the movies and series with a volume of 9 gigabytes. This dataset is available in both parts of products and reviews. In the products section, the title and price of the product have been indicated and the second section of this dataset which covers the reviews includes parts such as user ID, profile name, user score for the products and user comments on them. The most important part of this dataset is the users’ rating and their comments on the products. The role of these two items is very significant in this study. The final recommender system works by the combination of two parts which one of them is about the users’ comments and the other one is about the user rating.

Due to the large volume of the available dataset, here we have used Spark processor in Hadoop to implement and process data and recommend the users in a fast time. To implement it, firstly we process the primary dataset and create a new one which has a new record and

also the users' comments on the movies. Our recommended method performs its recommendations in two steps; first a typical recommendation and second by using the users' comments on the movies; the architecture of the suggested method in Spark is shown in the figure 2.

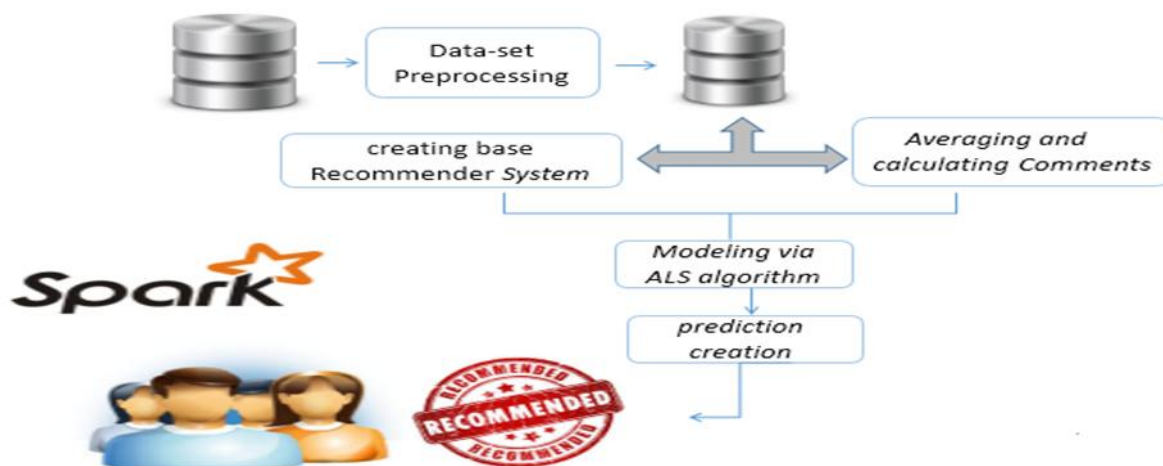


Fig.2. Application of spark platform in the proposed system

The following criterion with an appropriate ability to investigate the performance, model's accuracy and the suggested method has been used.

$$\text{precision} = \frac{TP}{TP + FP} \quad ; \quad \text{Formula 1}$$

Where TP indicates the number of items the user is interested in and have been recommended to him right, FP is the number of items the user is not interested in and have been suggested to him/ her by mistake. In this criterion, it is said that how many percent of the recommended items to the users are right.

4. Experiments and results

To evaluate the system, we compared the results obtained from the previous recommender system; in table 1, the neighborhood results for all users in dataset are investigated that better results have been achieved compared to the typical methods.

Table 1. Results for all users in dataset

	Precision in top 3 (%)	Precision in top 5 (%)	Precision in top 7 (%)
Standard CF	79/7	68/7	69/3
Proposed method	69/9	70/1	70/3

And in the second evaluation, the neighborhood results for 3 users existing in the dataset are investigated in table 2 and when the top list is 7, better results are obtained.

Table 2. The neighborhood results for 3 users

	Precision in top 3 (%)	Precision in top 5 (%)	Precision in top 7 (%)
Standard CF	64/3	64/3	62/7
Proposed method	64/6	64/8	63/9

5. CONCLUSION

In this study, we explored the users' comments and analyzed sentiments in order to improve the results obtained from the recommender system. In the survey stage, the unmonitored method based on Sentiwordnet dictionary has been used to calculate polarization. Due to limited used dataset in this research, sentiments have been considered in documental level; so that it obtains the general sentiments from the sentences and expresses the final sentiment about the item. The Spark processor in Hadoop context has been used due to the large data volume and in order to increase the rate and time of recommendation. In the previous researches, a few works have been done in the field of combining survey and the

recommender systems; but in these works, the large data volume has not be considered; in this study, we have considered the large data volume and obtained more acceptable results compared to the previous works.

In future, to improve recommendations to the users, this work can be done in the aspects based on the features of items and products, so that the users' preferences are obtained more exactly and they will be given more timely recommendations.

6. REFERENCES

- [1] Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook (pp. 1-35). Springer US.
- [2] Zainuddin, N., & Selamat, A. (2014 September). Sentiment analysis using Support Vector Machine. In Computer Communications and Control Technology (I4CT), 2014 International Conference on (pp. 333-337), IEEE.
- [3] Karau, H., Konwinski, A., Wendell, P., & Zaharia, M. (2015). Learning Spark: Lightning-Fast Big Data Analysis. "O ' Reilly Media, Inc."
- [4] Rosa, R. L., Rodriguez, D. Z., & Bressan, G. (2015, January). Music recommendation system based on user's sentiments extracted from social networks. In Consumer Electronics (ICCE) 2015 IEEE International Conference on (pp. 383-384), IEEE.
- [5] Siroos Keshavarz, Ali Reza Honarvar, (2015), A Parallel Paper recommender system in Big Data Scholarly, International Conference on Electrical Engineering and Computer.
- [6] Leung, C. W., Chan, S. C., & Chung, F. L. (2006, August). Integrating collaborative filtering and sentiment analysis: A rating inference approach. In Proceedings of the ECAI 2006 workshop on recommender systems (pp. 62-66).
- [7] Singh, V. K., Mukherjee, M., & Mehta, G. K. (2011). Combining collaborative filtering and sentiment classification for improved movie recommendations. In Multi-disciplinary Trends in Artificial Intelligence (pp. 38-50). Springer Berlin Heidelberg.
- [8] Kurian, D. D. M. K., Vishnupriya, S., Ramesh, R., Divya, G., & Divya, D. Big Data Sentiment Analysis using Hadoop. International Journal for Innovative Research in Science and Technology, 2015, 1(11), 92-96.

-
- [9] Ma, M. M. & Wang, S. P. (2015, June). Research of User-Based Collaborative Filtering Recommendation Algorithm Based on Hadoop. In International Conference on Computer Information Systems and Industrial Applications. Atlantis Press.
- [10] Maas, A. L., Daly, R. E. Pham, P. T. Huang, D. Ng, A. Y. & Potts, C. (2011 June). Learning word vectors for sentiment analysis. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1 (pp. 142-150). Association for Computational Linguistics.
- [11] Tian, P., Zhu, Z., Xiong, L., & Xu, F. A recommendation mechanism for web publishing based on sentiment analysis of microblog. *Wuhan University Journal of Natural Sciences*, 2015, 20(2), 146-152

How to cite this article:

Esfandiari Kh, Honarvar A, Aghamirzadeh AR. Improvement of recommender systems considering big data of users' comments on chosen items. *J. Fundam. Appl. Sci.*, 2016, 8(2S), 782-891.