ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

A Survey on Sentiment Analysis Based Keyword Aware Service Recommendation for Big Data

Khushboo R. Shrote, Prof A.V.Deorankar

M.Tech Student, Dept. of CSE, Government College of Engineering, Amravati, India

Associate Professor, Dept. of I.T., Government College of Engineering, Amravati, India

ABSTRACT: Recommender systems are the systems that can give personalized recommendations or can suggest best service among all when a number of options are available. Traditional Recommender systems suffer from scalability and inefficiency. Moreover most of the existing recommender services presents same results as the ratings given. So, a keyword aware service recommendation method is suggested in this paper. KASR gives personalized recommendation. Both active users preferences and passive users reviews and sentiments in the text are considered for score calculation. Sentiment Analysis is applied on these reviews to provide more accuracy. This method will be implemented on Hadoop using MapReduce paradigm. Hadoop will significantly improve the accuracy and efficiency by providing data set clustering across multiple nodes and fault tolerance in case of system failure.

KEYWORDS: recommender systems, Hadoop, MapReduce, keywords, preferences, Sentiment Analysis.

I. INTRODUCTION

Big Data analysis is one of the important tasks in Data Mining. When the amount of data that needs to be processed is greater than the capacity of the system it is known as Big Data. The three characteristics of big data are volume, velocity and variety. Volume refers to the amount of data that needs to be processed. It has moved to zetabytes. Velocity refers to the speed at which data can be processed with minimal error rate. Variety refers to all types of data such as unstructured, semistructured and structured data. Unstructured data is the data in the raw form such as facebook and twitter logs, text files, word documents etc. This data can not be easily analyzed. Semistructured data is in the form of xml files. Structured data is the data which has a well defined schema. It is in the form of rows and columns and can be easily retrieved using SQL.

This exponential growth of data has led to many challenges in IT industry and Businesses. Existing tools are inadequate to analyse and process such large amount of data. In order to overcome this Google introduces MapReduce algorithm. This algorithm divides the task into small parts and assigns them to many computers and collects the result from them which when integrated, form the result dataset. Using the solution provided by Google, Daug Cutting and his team developed an open source project called Hadoop. Hadoop runs applications using the MapReduce algorithm where data is processed in parallel with others. It is used to develope applications that could perform complete statistical analysis on huge amount of data. There are many big data analysis tools but the features that made hadoop distinct from others are :

Accessible : Hadoop can run on distributed clusters of nodes as commodity hardwares or on some services of cloud computing

Robust : Hadoop is designed to tolerate hardware malfunctions such as shut down or data loss. Three copies of same data are stored in different nodes. It can handle most of the failures with the secondary namenode.

Scalable : Once the multinode cluster is has been set up,more nodes can be added dynamically.

Simple : With the help of Hadoop users can easily write parallel codes.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

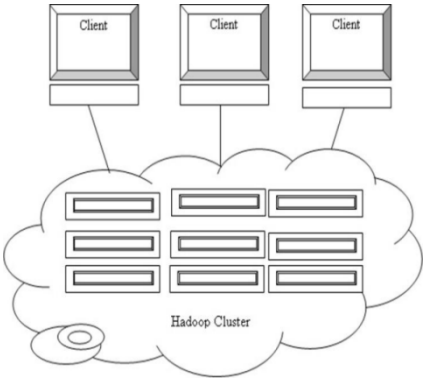

Fig 1.1. Multinode Clusters

MapReduce is a programming model where large sets of data can be distributed among the nodes of a cluster and processed parallel. There are two types of node such as Master node or name node and Slave node or data node . Master node is master of the system. It maintains and manages the blocks which are present on the slave node. It stores the metadata. Slave nodes are the nodes which are deployed on each machine and provide the actual storage. It is responsible for serving read and write requwsts for the clients. This model has two main steps which are 1) Map – Distribute the job among the slaves and 2) Reduce – Collect the results.

Sentiment Analysis is the software that captures feelings, attitudes, emotions and opinions. It allows businesses to track new product perception, brand perception, reputation management and flame detection. It allow individuals to get opinion on global scale.

II. LITERATURE REVIEW

Existing recommendation systems recommend services according to ratings given to that service. User can give only one keyword as a preference.

Shunmei Meng in 2014 proposed a KASR method for personalized recommendation. In this user based collaborative filtering is used. For more efficiency the method is implemented on Hadoop. For evaluation Jaccard coefficient and Cosine similarity measure is used. Users positive and negative reviews are not considered separately. Sentiments in the reviews are not considered.

Xiwang yang in 2013 proposed a Bayesian inference based recommendation in online social networks. In this users share their content ratings with friends. Rating similarity is measured using conditional probability. Based on similarity recommendation is done. Cold start and rating sparseness problem.

Guosheng kang in 2012 proposed an active web service recommendation. Web usage history and QoS are the main criteria. Using this recommendation top k services are generated for users. Passive users reviews are not considered.

Yan Ying Chen in 2013 proposed a probabilistic personalized travel recommendation model. People attributes and photos are used which are effective for mining demographics for travel landmarks and paths, and thus greatly benefiting personalized travel recommendation. Need to expand reaserch work to include more attributes for accuracy.

Faustuno Sanchez in 2012 proposed a recommender system for sport videos, transmitted over the Internet and/or broadcast, in the context of large-scale events, which has been tested for the Olympic Games. The recommender is based on audiovisual consumption and does not depend on the number of users, running only on the client side. This avoids the concurrence, computation and privacy problems of central server approaches in scenarios with a large number of users, such as the Olympic Games. Specific video fragment can not be recommended using this approach.

Zibin Zheng in 2013 proposed quality of service ranking prediction for cloud services. This paper investigate the combination of rating based approaches and ranking based approaches, so that the users can obtain QoS ranking prediction as well as detailed QoS value prediction. how to detect and exclude malicious QoS values provided by users is not proposed here.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 10, October 2015

III. CONCLUSION AND FUTURE WORK

In all above papers only one keyword can be used for preference. Old mathematical concepts of conditional probability are used. Passive users reviews are not differentiated into positive and negative reviews. Sentiments in the text are not considered. In Active Web Service recommendation usage history and QoS are mainly considered. But still there is scope in this area. We can use concept of Sentiment Analysis for more efficiency and accuracy. Research can also be done in the area where a different term appears than the domain thesaurus. More specifically,a keyword candidate list and domain thesaurus can be provided to help obtain users' preferences. The active user can give his/her preferences by selecting the keywords from the keyword-candidate list, and the preferences of the previous users can be extracted from their reviews for services according to the keyword-candidate list and domain thesaurus. These Reviews are further classified into positive reviews and negative reviews. A sentiment analysis based approach can be used to capture sentiments and emotions in reviews. A score corresponding to these reviews is calculated and high score service will be recommended first.

REFERENCES

1. Shunmei Meng, Wanchun Dou, Xuyun Zhang, Jinjun Chen," KASR: A Keyword-Aware Service Recommendation Method on MapReduce for Big Data Applications" IEEE Transactions On Parallel And Distributed Systems, TPDS-2013-12-1141
2. X. Yang, Y. Guo, Y. Liu, "Bayesian-inference based recommendation in online social networks," IEEE Transactions on Parallel and Distributed Systems, Vol. 24, No. 4, pp. 642-651, 2013.
3. G.Kang, J. Liu, M. Tang, X. Liu and B. cao, "AWSR: Active Web Service Recommendation Based on Usage History," 2012 IEEE 19th International Conference on Web Services (ICWS), pp. 186-193, 2012.
4. Yan-Ying Chen, An-Jung Cheng, "Travel Recommendation by Mining People Attributes and Travel Group Types From Community-Contributed Photos" IEEE Transactions on Multimedia, Vol. 15, No. 6, October 2013.
5. M. Alduan, F. Alvarez, J. Menendez, and O. Baez, "Recommender System for Sport Videos Based on User Audiovisual Consumption," IEEE Transactions on Multimedia, Vol. 14, No.6, pp. 1546-1557, 2013.
6. Zibin Zheng, Xinmiao Wu, Yilei Zhang,Michael R. Lyu, Fellow,and Jianmin Wang," QoS Ranking Prediction for Cloud Services" IEEE Transactions On Parallel And Distributed Systems, Vol. 24, No. 6, June 2013.
7. G. Linden, B. Smith, and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering," IEEE Internet Computing, Vol. 7, No.1, pp. 76-80, 2003.
8. Fuzhi Zhang, Huilin Liu, Jinbo Chao, "A Two-stage Recommendation Algorithm Based on K-means Clustering In Mobile E-commerce", Journal of Computational Information
Systems, Vol. 6, Issue 10, pp. 3327-3334, 2010.
9. Brian McFee, Luke Barrington and Gert Lanckriet, "Learning Content Similarity for Music Recommendation" IEEE Transactions on Audio, Speech, and Language Processing, Vol. 20, No. 8, 2012.
10. Z. D. Zhao, and M. S. Shang, "User-Based Collaborative-Filtering Recommendation Algorithms on Hadoop," In the third Internation-al Workshop on Knowledge Discovery and Data Mining, pp. 478-481, 2010.