ISSN:0975-9646

A Cluster-Based Filtering Algorithm for Mashup Service

Sathish.S,
Student,IFET College of Engineering,
Tamil Nadu,India

Nulyn Punitha J,M.Tech.,
Senior Assistant Professor,
IFET College of Engineering
Tamil Nadu,India

***Abstract*-- This system proposes a semantic analysis on the description of the text. The Mash up provides a flexible and easy-of-use way for service composition on web. Mash up is an ad hoc composition technology of Web applications that allows users to draw upon content retrieved from external data sources to create value-added services. Then, services are merged into clusters according to their characteristic similarities. To users, mining their implicit interests from usage records or reviews may be a complement to the explicit interests (ratings). as the ratings of services in the same cluster are more relevant with each other than with the ones in other clusters. Recommendations can be generated even if there are few ratings. This tends to solve the sparsity problem.**

Keywords—cluster; mashup; rating

I. NTRODUCTION

Big data is an all-encompassing term for any collection of data sets so large and complex that it becomes difficult to process using traditional data processing applications. The challenges include analysis, capture, duration, search, sharing, storage, transfer, visualization, and privacy violations. The trend to larger data sets is due to the additional information derivable from analysis of a single large set of related data, as compared to separate smaller sets.Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time. Big Data concerns large-volume, complex, growing data sets with multiple, autonomous sources. With the fast development of networking, data storage, and the data collection capacity, Big Data are now rapidly expanding in all science and engineering domains, including physical, biological and biomedical sciences. . For major Big Data-related applications, such as Google, Flicker, Facebook, and Walmart, a large number of server farms are deployed all over the world to ensure nonstop services and quick responses for local markets. Such autonomous sources are not only the solutions of the technical designs, but also the results of the legislation and the regulation rules in different countries/regions. Big data is difficult to work with using most relational database management systems.

Fig.no 1 Big data

II. RECOMMENDER SYSTEMS

Recently, recommendation systems are attracting more and more attentions, because it can help users to deal with information overload, which is a great challenge in the modern society, especially under the exponential growth of the Internet and the World-Wide-Web. There has been much work done both in the industry and academia on developing new approaches to recommender systems over the last decade. The interest in this area still remains high because it constitutes a problem-rich research area and because of the abundance of practical applications that help users to deal with information overloads and provide personalized recommendations, content, and services to them. Users offer feedback on purchased or consumed items, and the recommender system uses the information to predict their preferences for yet unseen items and subsequently recommends items with the highest predicted relevance. Recommendation algorithm has been used to recommend books and CDs at Amazon.com, movies at named Netflix.com,and newsatversify Technologies (formerly AdaptiveInfo.com).

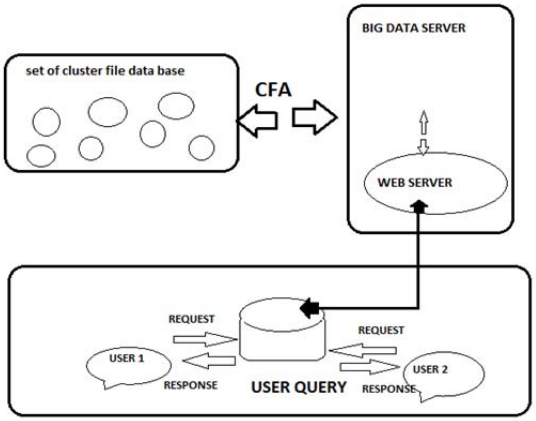

Fig.no 2 Architecture

III. MAP REDUCE

A Map Reduce program is composed of a Map procedure that performs filtering and sorting and a Reduce procedure that performs a summary operation. The "Map Reduce System" orchestrates the processing by marshaling the distributed servers, running the various tasks in parallel, managing all communications and data transfers between the various parts of the system, and

"Map" step: Each worker nodes applies the "map ()" function to the local data, and writes the output to a temporary storage. A master node orchestrates that for redundant copies of input data, only one is processed.

"Shuffle" step: Worker nodes redistribute data based on the output keys (produced by the "map ()" function), such that all data belonging to one key is located on the same worker node.

"Reduce" step: Worker nodes now process each group of output data, per key, in parallel.

IV. CONCLUSION

Here we proposed cluster based algorthim for mash up service. Service users have nowadays encounter unprecedented difficulties in finding ideal ones from the overwhelming services. The service-relevant data become too big to be effectively processed by traditional approaches. To users, mining their implicit interests from usage records or reviews may be a complement to the explicit interests. Service users have nowadays encounter unprecedented difficulties in finding ideal ones from the overwhelming services. The service-relevant data become too big to be effectively processed by traditional approaches To users, mining their implicit interests from usage records or reviews may be a complement to the explicit interest. A naïve solution is to decrease the number of services that need to be processed in real time.

REFERENCES

[1] M. A. Beyer and D. Laney, "The importance of ″big data″: A definition," Gartner, Tech. Rep., 2012.

[2] X. Wu, X. Zhu, G. Q. Wu, et al., "Data mining with big data," IEEE Trans. on Knowledge and Data Engineering, vol. 26, no. 1, pp. 97-107, January 2014.

[3] A. Rajaraman and J. D. Ullman, "Mining of massive datasets," Cambridge University Press, 2012.

[4] Z. Zheng, J. Zhu, M. R. Lyu. "Service-generated Big Data and Big Data-as-a-Service: An Overview," in Proc. IEEE BigData, pp. 403-410, October 2013.

[5] A. Bellogín, I. Cantador, F. Díez, et al., "An empirical comparison of social, collaborative filtering, and hybrid recommenders," ACM Trans. on Intelligent Systems and Technology, vol. 4, no. 1, pp. 1-37, January 2013.

[6] W. Zeng, M. S. Shang, Q. M. Zhang, et al., "Can Dissimilar Users Contribute to Accuracy and Diversity of Personalized Recommendation?," International Journal of Modern Physics C, vol. 21, no. 10, pp. 1217-1227, June 2010.

[7] T. C. Havens, J. C. Bezdek, C. Leckie, L. O. Hall, and M. Palaniswami, "Fuzzy c-Means Algorithms for Very Large Data," IEEE Trans. on Fuzzy Systems, vol. 20, no. 6, pp. 1130-1146, December 2012.

[8] Z. Liu, P. Li, Y. Zheng, et al., "Clustering to find exemplar terms for keyphrase extraction," in Proc. 2009 Conf. on Empirical Methods in Natural Language Processing, pp. 257-266, May 2009.

[9] X. Liu, G. Huang, and H. Mei, "Discovering homogeneous web service community in the user-centric web environment," IEEE Trans. on Services Computing, vol. 2, no. 2, pp. 167-181, April-June 2009.

[10] H. H. Li, X. Y. Du, and X. Tian, "A review-based reputation evaluation approach for Web services," Journal of Computer science and technology, vol. 24, no. 5, pp. 893-900, September 2009.