HAL
open science

Context Aware Service Information Extraction from Bigdata

G Suvarna, Krihna Murali

To cite this version:

G Suvarna, Krihna Murali. Context Aware Service Information Extraction from Bigdata. International Journal of Computer Science and Information Technologies, 2015. hal-01267127

HAL Id: hal-01267127

https://inria.hal.science/hal-01267127v1

Submitted on 22 Apr 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ISSN:0975-9646

G.Suvarna et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (1) , 2015, 368-373

Context Aware Service Information Extraction from Bigdata

G.Suvarna
(M.Tech) Department of Computer Science & Engineering,
Madanapalle Institute of Technology &Sciences,
Madanapalle, Andhra Pradesh, India

Dr. S. Murali Krishna
Professor and Head of the Department, Department of Computer Science & Engineering,
Madanapalle Institute of Technology &Sciences,
Madanapalle, Andhra Pradesh, India.

Abstract: **Big data is a term used for defining the exponential growth and availability of both structured and unstructured and semi structured data is most difficult for processing with traditional database systems. Big data is inward from multiple sources at an alarming velocity, variety and volume. Although big data does not refer any specific quantity the term is used when speaking for pet bytes and Exabyte's of data. In existing method some keywords are used for indicating the user preferences and they used user based collaborative filtering algorithm for generating services to the users. They did not consider about the user preferences. In this paper we proposed a context aware approach for addressing the above challenges. The main aim of this paper is to make the users personalized recommendation for recommending the appropriate services to the users. Context aware collaborative filtering algorithm is used for meeting the requirements of users.**

Keywords: **Context awareness, context aware recommendation, decision support, recommendation system, big data.**

I. INTRODUCTION

Nowadays, the world may annotate with petabytes of electronic system also service details dispensed around various available global information databases. People of many methods as well as services are consistently distributing further details or giving the information databases with their choices and experience. Smartphone's as well as tablets operate as a window for monitoring and obtaining these annotated details and to render the consumers an input unit being accumulates further details. E-business, promoting and e-commerce is benefiting a lot by this persistent use of alternative device also customer data. Global industries, like eBay, Amazon, Apple iTunes or Google Play, provide millions of countless items as well as services in hundreds of types. These types cross a wide range of product groups from ordinary hardware to software also mobile applications.

EBooks, gadgets, video and music streaming or even food. The large quantity of consistent obtainable details renders it complicated or even not possible. For consumers to manually choose an appropriate subset. As an individual user is not inclined to analyze all accessible data, the choice of that subset is of essential value for both, the individual consumer and for the data writers. The most prevalent real world situation is a individual user exploring for a system or service also a large amount of providers supplying data on their particular offer.

Big Data is about growing challenge that organizations faces as they deal with large and fast growing sources of data or information that also present a complex range of analysis. Effective analysis of Bigdata provides a lot of business advantage as organizations will learn which areas to focus on and which areas are less important. Big Data analysis provides some early key indicators that can prevent the company from a huge loss or help in grasping a great opportunity with open hands. A precise analysis of Big Data helps in decision making. For instance, now a day's people rely so much on Face book or twitter before buying any product or service. With time, data volume is growing exponentially. Earlier we used to talk about Megabytes or Gigabytes. But time has arrived when we talk about data volume in terms of terabytes, petabytes and also zettabytes! To address that issue, global industry eventually distinguished the demand for trustworthy item suggestion inside their techniques. In 2006, the Netflix Prize challenge was started with a 1 million dollar prize for obtaining a ten percent or more upgrade of Netflix's video referral algorithm. The knowledge specify that Netflix revealed for the price challenge included about 100 million rankings from around 500.000 incognito clients on 17.000 videos. The competition had drawn 48.000 rivaling groups from 182 varied nations. The receiving group (BellKor) from AT&T Research laboratories (contains Bob Bell as well as Chris Volinsky, from the Statistics Research team in AT&T Labs, also Yehuda Koren) was ready to enhance the efficiency of Netflix's referral algorithm by 8.43 percent. So it's noticeable that ordinary referral techniques perform a vital role in modem individual industries. When referral techniques for ordinary product referral, like Slope One referral or Matrix Factorization, have been commonly revealed inside the past ten years, most intriguing elements of client- centric referral techniques haven't been inside the emphasis by the referral data network yet. Bell et al. determined numerous such data items through their focus on the Netflix award challenge.

II. Review Of Related Research

In recent years the usage of data in our world has been increased from megabytes to zetta bytes. One way of looking at big data is that it represents the large and rapidly growing volume of information that is mostly untapped by existing analytical applications and data warehousing systems. Big data is a useful term it is a new kind of data management and analysis of large volumes of data sets whose size is beyond the capability of current technology. Now big data is a marketing buzzword used to depict large amounts of both structured and unstructured and semi structured data that is so big and difficult to process the data using a traditional database and software techniques.

Similar to most applications of big data, the big data trend also impacts on service recommender systems. Service recommender systems are the recommender systems which provide services to the users based on the user preferences, it provide appropriate recommendations to the user. With the increasing number of different services, so recommending the effective service to the user is becomes an essential research issue. Examples of such recommender systems are recommending the books, cd's and various products those are needed by the user. Over the last years, there has been a research done in both industry and academia for finding the new approaches for service recommender systems. Recommender systems are becoming an significant research area since the mid-1990s there has been a large amount of work done in the industry and academia for developing new methods to recommender systems. The significance of context-awareness in human-centered processing techniques has been reviewed by many assorted analysis forums, such as ubiquitous and persistent processing, mobile processing, e-commerce as well as e-business, data recovery as well as selection, promoting as well as administration and inside numerous engineering procedures. During the significant improve of hardware possibilities in blend with affordable broadband connection in customer technology, like mobile phones as well as tablet PCs, the desire for context-related data selection is significantly elevating also. Recommendation algorithms are best for using on e-commerce websites where they are using customer interest as input for generating a list of recommended items.

Applications may use the items that customer purchase and they give a rate to represent their interests but they can also use other parameters like items viewed, demographic data, favorite artists and subject interests. We use recommendation algorithms for personalizing the online store to each customer at amazon.com.The store may change based upon the customer interests like showing programming title to the software programmer , related books for the students and baby toys to a new mother-commerce recommendation algorithms are running in challenging environment for example: A large retailer might have long volumes of data and millions of customers and distinct catalogue items and another example is new customers have limited information and old customer has glut of information from the previous purchased items and catalogus.There are three approaches for solving the recommendation problems they are by using collaborative Filtering, cluster models, and search-based methods.

Broadened the extent of context data to temperature, time, season also various aspects. Because essentially the many context data aspects is inexhaustible, the classification of context by Anhind K. Dey is the most frequently used: Context is any data that can be employed to define the scenario of an entity. An entity is a person, place, or object that is regarded appropriate to the connection concerning a user as well as an application, such as the consumer and application independently [5]. This description of context determines that context consists of whichever data regarding an entity to comprehend its scenario. Frequently the label context is constrained to location data as well as location-awareness, but in modern year's context also is overflowing with the social community of a consumer. Obtaining and assessing the social aspect of context associated to a particular user proceeds together with an illustrated conversation about privacy as well as security. An exciting concept about the above description of context is that Dey determines three base classes that categorize all objects: person, place and object. This type of definition has pragmatic motives but is even addressed to a location-dependent perspective of context data. Through the last decade various architectures as well as achievements of system middleware frameworks have been released that highlighted the aggregation and understanding of context-information. Recommender systems use user preferences for suggesting item to purchase or for examining the items. They have more importance in applications like electronic commerce and information access and they provide suggestions for pruning the huge amount of information spaces so that a user can easily directed towards the items which they meet their needs .

Different techniques are used for developing the recommendations including content based, collaborative and knowledge based techniques. for improving the performance they combined these techniques with hybrid recommender systems. In this work they use both knowledge based and collaborative filtering for recommending services in restaurants .Recommender systems are defined as the recommendations provided by user as input and then the system aggregates it provides for appropriate recipients. Recommender systems must have the background data that is the data presented before the recommendation process begins, input data that is for generating recommendations the user must communicate with the system and an algorithm which combines both input data and background data for arriving its suggestions. This work surveys various recommendation techniques are being researched and analyzing the results on data which supports the recommendations and algorithms that are operated on data and it views the hybridization techniques that have proposed.

The essential idea regarding a lot of the research strategies inside context-acquisition, handling and understanding is to utilize a customer's context data being separate important data (e.g., on products, services, locations) from the large collection of supplied data. The authors Shunmei Meng, Wanchun Dou, Xuyun Zhang and Jinjun Chen proposed a

keyword-aware service recommendation method on map reduce for big data applications. It aims at presenting a personalized service recommendation list and recommending the most appropriate services to the user effectively. Specifically keywords are used to indicate users' preferences, and a user-based Collaborative Filtering algorithm is adopted to generate appropriate recommendations.

The disadvantages of existing system are about the accuracy and they did not consider the negative preferences of users. With the motivation gained from KASR, we will do further research in how to deal with the case where term appears in different categories of a domain thesaurus from context and how to distinguish the positive and negative preferences of the users from their reviews to make the predictions more accurate. Often the term context is limited to location information and location-awareness, but in recent years context also is enriched with the social network of a user. Collecting and evaluating the social dimension of context related to a specific user goes hand in hand with a detailed discussion about privacy and security. The significance of context-awareness in human-centered processing techniques has been reviewed by many assorted analysis forums, such as ubiquitous and persistent processing, mobile processing, e-commerce as well as e-business, data recovery as well as selection, promoting as well as administration and inside numerous engineering procedures. During the significant improve of hardware possibilities in blend with affordable broadband connection in customer technology, like mobile phones as well as tablet PCs, the desire for context-related data selection is significantly elevating.

An strategy that attempts to address the equivalent concern is to utilize recommendation techniques and algorithms to choose a subset of content that tends to be pertinent for a consumer. Regarding referral methods have long custom inside global industry. Ordinary referral techniques accept a set U of consumers and a group of products (objects) P, which must be suggested to a consumer. A referral strategy then offers a feature function f that determines the significance of an item from set P to a provided consumer. This service function f ($f : U \times P \rightarrow s$ R, where R is an ordered group of numbers) assigns a ranking to every product (or equal to a composite group of products) in a strategy that catches the significance or inclination for a particular customer. The aim of referral techniques is to choose or determine this service function f. Function f is applied to estimate the significance of products from P and of newer appearing products with equivalent elements. In the literature various strategies occur for choosing a function *f* by utilizing an obtainable dataset. Ordinary referral techniques are recognized into two major techniques: content filtering as well as collaborative filtering.

A).Content Filtering

This perspective produces information for every product and consumer, in order to define and evaluate its quality. Every profile includes a certain group of characteristics, which can be utilized to evaluate items. For instance, a restaurant could have a delicacies attribute, explaining the kind of food it supplies, a location feature, a vegetarian label, etc. A referral function f selects products that are equivalent to products the customer has already selected or ranked earlier. The service function analyzes the customer's profile as well as determines the resemblance of a consumer profile with the supplied products. So, the customer profile enables the referral system to produce a set of products that could blend to a provided customer profile. Most achievements of this strategy further pertain to Linked Data content, like RDF stores as well as Semantic Web databases, to categorize and search consistently for associated data.

B).Collaborative Filtering

These perspectives, the referral function selects products that were recommended by other customers with same characteristics. Collaborative filtering approaches rely on either specific or implied user rankings of products. By ranking various products a customer can supply specific rankings into the referral system, while implied feedback is accumulated by the method with the evaluation of the customers tendencies (earlier products, routing path, search provisions, and so on.). Collaborative filtering is domain-free, which indicates that it can be used to any program location and to various data items, which could be difficult to produce into a specific profile. Collaborative filtering is most precise than content filtering [7] but has the concern of beginning without any primary data sets (cold start issue). It is not exclusively feasible to manage new customers or items where the strategy has no primary data set provided. Prominent collaborative filtering techniques are district strategies and hidden factor designs. The Pearson's correlation coefficient sim(u, v) is frequently applied to determine the prominent region technique k Nearest Neighbor, to determine the resemblance around the target user u, and a neighbor v. Inside the Pearson's correlation the signal ru refers to a medium ranking of user u as well as P signifies the group of items or products.

One more strategy employs association rules to clearly design the resilience and resemblance of products In this work they analyze various item based recommendation algorithm generations they look into different methods for computing item-item similarities and various techniques for getting the recommendations from them some experiments suggesting that the item based algorithms are giving better performance than the user based algorithms. Normally the collaborative filtering gives its performance by building database with some preferences of user about some items for example a new user Alex that is matched against the database for finding the neighbors which are other users who are having the very similar taste to Alex. The items which are recommended by the neighbor are recommended to the Alex as he probably like this one. There are some important research questions to overcome the challenges of collaborative filtering. The first challenge is for making the scalability of collaborative filtering algorithms. These algorithms are having the capability of searching the thousands of neighbors in real time but the demands of modern systems are to search millions of neighbors. The second challenge is for improving the

quality of the recommendation for the specified users. Users need recommendations that they help to find the items they will like. In this work they address the issues of recommender systems by the item based algorithm. Item based algorithms analyze the relationships between items first rather analyzing the users. The recommendations for users are computed based on the items they like that stimulate a customized query language to establish adaptable and modified referral queries according to multidimensional OLAP- cubes. Benefits were made by research associations that designed assorted program conditions for context-aware referral techniques, extending from referral of views inside the tourism site, restaurants or also people (e.g., glancee.com).

In commercial recommender systems we normally use collaborative filtering algorithms for personalizing the recommendation of users for improving the performance but the drawback is its scalability when we are having the huge amount of input data then the computational cost of collaborative filtering is very high. By using HADOOP MapReduce we can achieve the scalability on HADOOP platform we can efficiently do the large amount of big files compare with the large amount of small files. In this work they proposed how to develop the collaborative filtering using map reducing. But they feel difficult while developing so they came up an idea with the mapper function that is to make their recommendation we store the user ids for calculating some txt files. MapReduce framework will assign some user ids to mapper function for this the algorithm is divided into three phases. They are

- Portioning phase
- Mapper phase
- Reducer phase

In portioning phase they divide the ids into various numbers of files each user_id is stored in separate row this files are input for the mapper phase.

In mapper phase first it reads the line number as a key and user id as a value and it will calculates the difference between one user to another user.

In reducer phase in hadoop some reducers are generated and I will take the mapper output as a reducer input and it processes in the form of sort and shuffle and it gives the result.

III. Context-Aware Collaborative Filtering Algorithm:

With the motivation gained from KASR, we will do further research in how to deal with the case where term appears in different categories of a domain thesaurus from context and how to distinguish the positive and negative preferences of the users from their reviews to make the predictions more accurate. Often the term context is limited to location information and location-awareness, but in recent years context also is enriched with the social network of a user. Collecting and evaluating the social dimension of context related to a specific user goes hand in hand with a detailed discussion about privacy and security. We use an algorithm context aware collaborative filtering for recommending the appropriate services to the user.

Input:

The recommendation process initializes with the specification of the initial set of ratings which are either explicitly provided by the users or implicitly inferred by the system.

Output:

Display the services which are recommended by the users once these initial ratings are specified, a recommender system tries to estimate the rating function R

$$\mathbf{R: User \times Item \rightarrow Rating}$$

Rating is a totally ordered set user and Item are the domains of users and items.

Once the function R is estimated for the User×Itemspace then ratings are defined with the rating function

$$\mathbf{R: User \times Item \times Context \rightarrow Rating;}$$

Where User and Item are the domains of users and items, Rating is the domain of ratings, and Context specifies the contextual information associated with the application.

By considering the contextual information associated with the users and locations for recommending the services. It contains a set of words which has the intersection tags between the user and location.

To recommend appropriate services to the active user u_a, the system needs to record $R(u_a,s)$ the ratings of services that have been visited by u_a, and predict $P(u_a,s)$ the ratings of services that have not been visited by u_a. Our context-aware collaborative filtering algorithm has three steps:

(1) Record the ratings of visited services R (u,s). The ratings are explicitly rated by users or implicitly inferred by the system.

(2) To predict the active user's rating $P(u_a,s_i)$ of an unvisited service s_i, we compute the similarity of u_a and the users who have rated service s_i, and select a neighborhood Usi consisting of k users most similar to u_a.

(3) Combine neighbors' ratings to predict $P(u_a,s_i)$. After predicting u_a's ratings of all unvisited services, we can sort and recommend the top most services to u_a.

The standard software design for constructing a context-aware referral strategy is originated into a typical client-server architecture design. This client-server design employs a MapReduce development system, as it was definitely revealed in the basic strategy with some essential subsystems. As it is revealed in Figure 1, the server describes every essential subsystems for data use also third-party data access, customer connects for manual content collection and modification, and the context-sensitive recommender. Primary section of the server execution is the control of sore context data with certain customer as well as item information. Semantic information that is applied to determine the similarity and connection between provided customers or products is retained with related data databases. All alternative semantic information can be utilized by employing semantic web criteria as well as query languages, like RDF and SPARQL. The objective of obtaining these sources of semantic data is to get further item-based Resemblance dimensions that are applied in blend with standard collaborative filtering outcome. Additional sources of semantic data, like Facebook or Last.FM, are either exclusively brought in and replicated, or exclusively utilized through a described service platform.

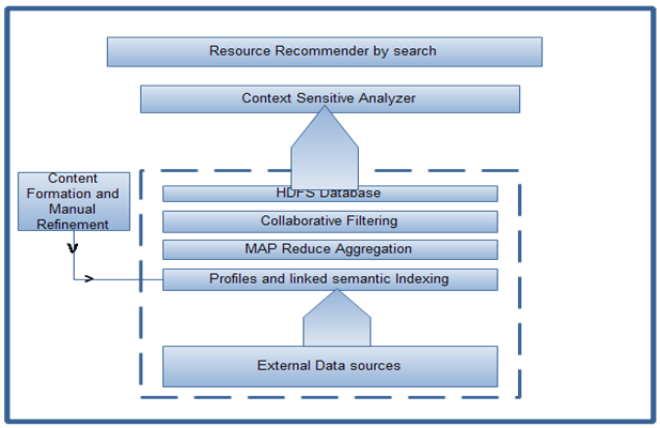

Fig i).General software architecture for the implementation of a context-aware recommendation system

Inside the map-reduce layer many specific map-reduce procedures are determining stabilized context-dimensions that are blended to a ordinary ranking table among customers and products. The MapReduce development design describes two basic steps: Map and Reduce. Through a synchronous Map step every dispensed listings accumulate the obtainable data designs. Inside this Map function every database models are accumulated that incorporate a rating between customers and products and produced as intermediate outcome.

IV. Experiments And Results

Inside this assessment segment certain illustrated outcomes regarding referral excellence and efficiency are envisioned. The excellence of referral outcomes inside this work is considered by determining a excellence rating by applying a regular absolute difference assessment technique (mean average error score) that splits the obtainable ranks into 70% learning data as well as 30% assessment data. The collaborative filtering technique applied inside this review is considering a matrix factorization representation of provided customers and event products onto an element place. Our assessment revealed that enhancing the quantity of attributes above a quantity of five does not considerably enhance the general referral rating. At this point of our assessment data, the assessment rating itself does not incorporate any effective data as many of the reviews are binary ranks conveniently removed from inactive user information. So the number of the ranks is between zero as well as one.

Along with the assessment of how the amount of properties enhances the referral score we even considered the efficiency of the entire computation. The efficiency checks were completed on a traditional Windows 7 Laptop product with 4GB RAM as well as a Intel Core i5 64bit CPU with 1.70GHz. The database comprises 408,634 inactive and effective customer information and 22,901 various events (products). Once the MapReduce development design aggregated the ratings, as it was revealed in Section V the subsequent quantity of ranks is 399,905.

The assessment of this structure reveals that it is an appropriate strategy and method for employing a context-sensitive referral system that utilizes a MapReduce development design in blend with collaborative filtering. The assessment of the excellence of referrals does not incorporate any considerable outcomes as the data set doesn't incorporate sufficient specific user ranks so far. The implementation model screen cast is explored by fig2, which followed by the graph representation of the information accuracy achieved due to the consideration of the context sensitivity as a factor (see fig 3). The fig 3 indicates that the proposed framework is accurate towards identifying the optimal results against to the search on HDFS data, and also is scalable when it compare to KASR.

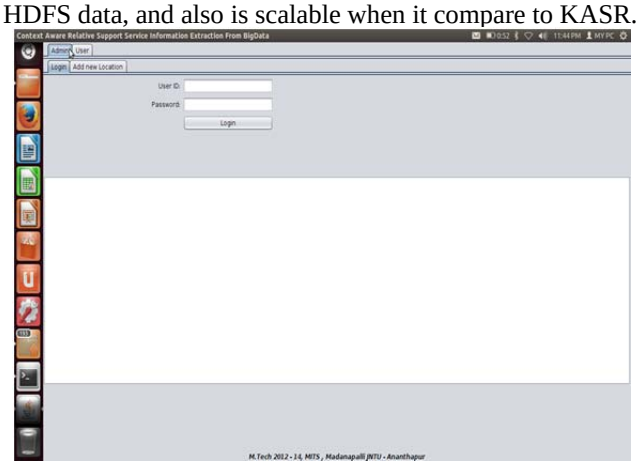

Fig a).Main screen to be used by admin to initialize

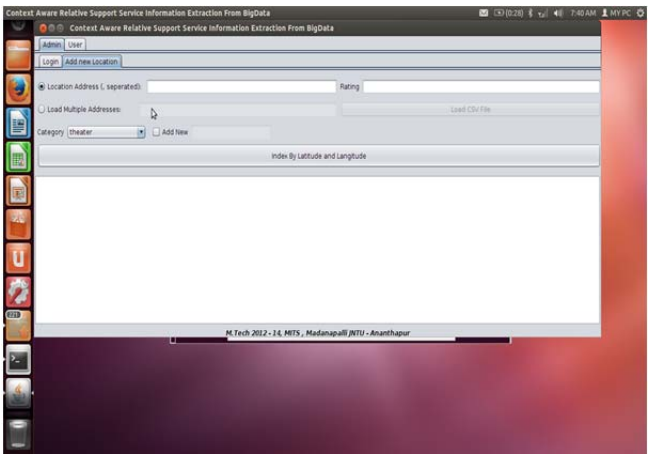

Fig b). Screen to configure a load data in csv format to index at HDFS

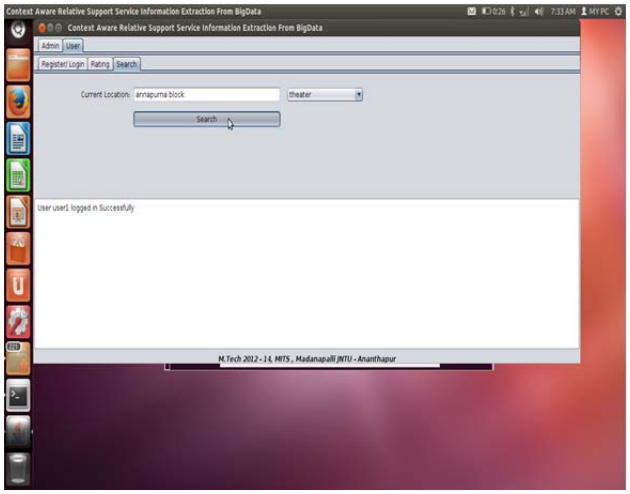

Fig c). User will search by giving current location

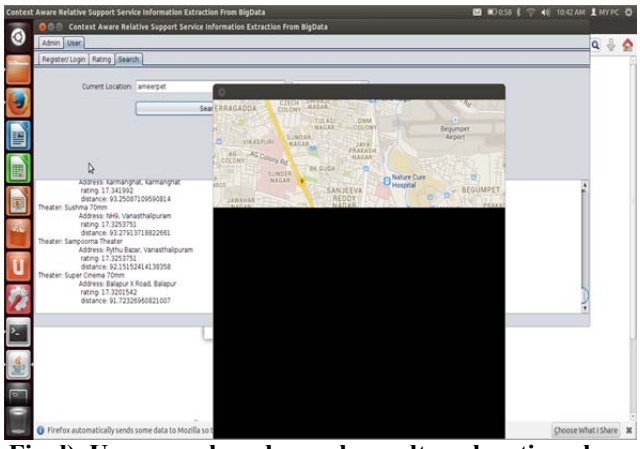

Fig d). User search and search result exploration along with map positioning of the target

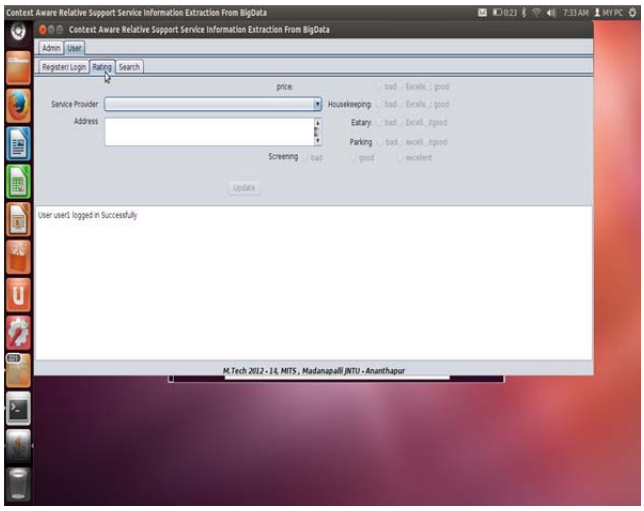

Fig e). User can provide ratings to the service

V. Conclusion

In that move, we recommend a normal MapReduce-based strategy, and software design for the execution of context-aware referral techniques. The strategy and the structure provides significant convenience based on the description and setup of newer context proportions in type of influence functions, which impact the referral of products for provided customers. The structure is domain-free, that indicates that this strategy can be employed and used for various software domains. The context-aware referral of products of each type that varying from items of e-commerce to strategies and expertise in recreation and fun gets a lot concern in upcoming software programming. A modification of a domain-specific referral system on top of our projected strategy could be applied with lowered developing work, as it is generally decreased to an easy choice of context proportions. We consider that a standard structure for generating and employing these referral techniques for various software domains is of excellent significance. The next steps inside our move will be to obtain scientific reviews from the group inside the provided use-case of suggesting music connected activities and to enhance the level of visibility for the referral strategy.

References

[1]. W. Beer, W. Hargassner, S. Herramhof, and C. Darwin, “General framework for context-aware recommendation of social events,” in Proceedings of the Second International Conference on Intelligent Systems and Applications (INTELLI). IARIA, 2013, pp. 141-146.

[2]. R. M. Bell, Y. Koren, and C. Volinsky, “The BellKor solution to the Netflix prize,” accessed: 31/01/2013.[Online].Available:http://www2.research.att.com/volinsky/netflix/ProgressPrize2007BellKorSolution.pdf

[3]. B. Schilit, N. Adams, and R. Want, “Context-aware computing applications,” in Mobile Computing Systems and Applications, 1994. WMCSA 1994. First Workshop on Mobile Computing Systems and Applications. IEEE, 1994, pp. 85-90.

[4]. 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.

[5]. A. Dey and G. Abowd, “Towards a better understanding of context and context-awareness,” in CHI 2000 Workshop on the What, Who, Where, When, and How of Context-Awareness, 2000.

[6]. W. Beer, V. Christian, A. Ferscha, and L. Mehrmann, “Modeling context-aware behavior by interpreted eca rules,” Euro-Par 2003 Parallel Processing, pp. 1064-1073, 2003.

[7]. R. Burke, “Hybrid Recommender Systems: Survey and Experiments,” User Modeling and User-Adapted Interaction, Vol. 12, No.4, pp. 331-370, 2002.

[8]. G. Adomavicius, and A. Tuzhilin, “Toward the Next Generation of Recommender Systems: A Survey of the State of- the-Art and Possible Extensions,” IEEE Transactions on Knowledge and Data Engineering, Vol.17, No.6 pp. 734-749, 2005

[9]. W. Beer and A. Wagner, “Smart books: adding context-awareness and interaction to electronic books,” in Proceedings of the 9th International Conference on Advances in Mobile Computing and Multimedia (MoMM). New York, NY, USA: ACM, 2011, pp. 218-222.

[10]. V.-G. Blanca, G.S. Gabriel and P.M. Rafael, “Effects of relevant contextual features in the performance of a restaurant recommender system,” in RecSys11: Workshop on Context Aware Recommender Systems (CARS-2011), 2011.

[11]. Y. Koren, “Collaborative filtering with temporal dynamics,” in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD). New York, Ny, USA: ACM, 2009, pp. 447-456.

[12]. N. Sundaresan, “Recommender systems at the long tail,” in Proceedings of the fifth ACM conference on Recommender systems (RecSys). New York, NY, USA: ACM, 2011, pp. 1-6.

[13] D. Bouneffouf: "Freshness-Aware Thompson Sampling", 21st International Conference Neural Information Processing ICONIP (3) 2014: 373-380