Journal of Engineering and Applied Sciences 13 (2): 523-528, 2018
ISSN: 1816-949X

A Hotel Recommendation System for Big Data Applications using a Keyword Aware Approach

Samiksha S. Bhujade and Manoj B. Chandak
Department of Computer Science and Engineering,
Ramdeobaba College of Engineering and Management, Nagpur, India

Abstract: A recommender system plays a very vital role in recommending services to the users. The users here can rate their preferences and choose the services according to their requirements. The different recommendation services include services like books, hotels, newspapers, etc. The online information today is increasing on a large scale the data is growing rapidly, yielding the problem of big data to the environment. So, to overcome the problem of scalability on such big data problems, the recommendation systems have taken steps to overcome the same. In this study, the hotel recommendation system is considered. In most traditional recommendation systems, the distinct preferences of the users are not considered, i.e., the identical list of hotels is provided to every user without considering the users distinct preferences. For this reason in this study the reviews of distinct users are considered, so that, users with the identical tastes can be provided similar hotels list and users with distinct tastes with different hotels list.

Key words: Recommender system, services, big data, hotel, preferences, newspapers

INTRODUCTION

The amount of data in the world is drastically increasing now a days. For analyzing and storing such large-scale data, the concept of big data comes into the picture here. The growth of productivity and the innovations done in the field of big data is worth viewing. Big data refers to the data sets whose size is beyond the capability of current technology to manage and process the data within an elapsed time.

This big data problem has a huge impact on the recommender systems. The recommender systems and its growing popularity with a large amount of data have yielded the service recommender systems into the big data environment. The recommender system plays a very vital role in helping the users to get the efficient services. Many practical applications of such recommender systems include services like books, hotels, newspapers, etc. In the last decade there has been much research done in developing new approaches for recommender system.

In most traditional hotel recommender systems, the services offered and the recommendation list provided for the user are the same. The recommendation list given to the users should be distinct for users with distinct preferences. So, in order to provide the recommendation list to the user, considering the preferences of the users is important. For this reason, the reviews of different users who have visited the hotel earlier are considered. If the reviews of the users contain the keywords which are similar to the requirements of the current user then the name of the hotel is recommended to the user. Finally, a list of top k hotels is recommended to the user assuming ratings and rankings (Fig. 1).

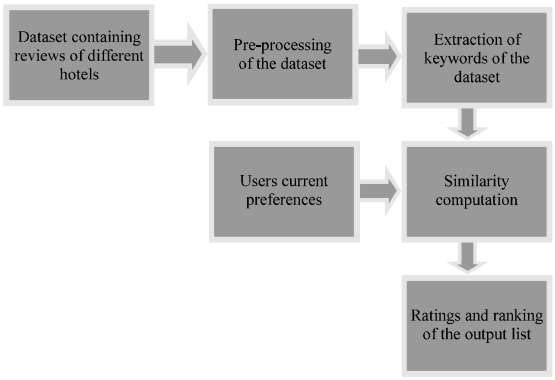

Fig. 1: Flowchart of the process

Motivation: With the emerge of Web 2.0, many companies started storing the big data information about their customers, providers and operations. For this reason there is a rapid and drastic growth of the number of

Corresponding Author: Samiksha S. Bhujade, Department of Computer Science and Engineering,
Ramdeobaba College of Engineering and Management, Nagpur, India

customers, services and other information had kept the service recommender systems in the 'big data' environment by which many critical challenges have been posed on the recommender systems. In most traditional recommender systems such as for the hotel reservation systems, the lists of recommendations provided to all the users are the same they are not considering the users different personal preferences. Following is an example of a hotel reservation system:

Example: Two persons Jack and Jones are respectively browsing through a hotel reservation website to reserve a hotel. But the recommendation lists provided to them are the same. Assuming that there are three hotels A-C. After comparing the three hotels, A is a posh hotel and has a free Wi-Fi facility. The rooms and food of hotel B are very good. And hotel C has good gym facility. Now, the preferences o f Jack and Jones are as follows: Jack prefers a hotel with a gym facility and also should have a free Wi-Fi. Whereas Jones prefers that the food at the hotel should be delicious.

Now, the problem is how to provide Jack and Jones the recommendations list according to their preferences because the website will provide both the customers the same list of hotels. Ideally, according to Jack's preferences hotel, A and C should be recommended to Jack and similarly according to Jone's preferences hotel B should be recommended to Jones. So, for this reason the reviews of the previous customers are taken into consideration. The reviews can be taken from the travel review site www.tripadvisor.com.

Motivated by these observations in this study, we address these challenges through the following contributions: A keyword aware approach for hotel recommendation system for big data applications is proposed in this study. The recommendation list will be provided for the customers according to their preferences.

Recommendation systems and collaborative filtering: As an independent research area, recommender systems started developing in the mid 1990's recommender systems can be defined as the systems which provide individualized recommendations list to the users according to their preferences. The recommender systems can be classified into three main categories: Content-based, collaborative and hybrid services of recommender systems. The content-based filtering is also referred to as cognitive filtering, it does a comparison between the content of the items and a user profile and recommends items based upon the comparison. The collaborative filtering is also referred to as social filtering by using the recommendations of other people it filters the information. It is based upon the idea that people who agreed in their evaluation of certain items over the past are likely to agree again in the future. The hybrid approach is the combination of the other two approaches.

Literature review: Many prior researchs have been done on the applications of recommender systems. Greg Linden, Brent Smith and Jeremy York at the Amazon.com use recommendation algorithms to personalize the online store for each customer. They used the concept of item-to item collaborative filtering (Linden *et al*., 2003).

Adomavicius and Kwon (2007) gives an improvement of understanding of users and items and the incorporation of the contextual information about the recommendation process.

Bjelica (2010) has proposed a TV recommendation system. He emphasizes on user modeling algorithms that would be able to efficiently learn the user's interest.

Sanchez *et al*. (2012) described 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.

Asawarangsee and Maneeroj (2003) have proposed a technique for multi-criteria recommendation. Gobin and Subramanian had proposed knowledge modeling for a hotel recommender system in this study a methodology for the development of the knowledge model which is inspired by both software and knowledge engineering is proposed (Gobin and Subramanian, 2007).

Takuma *et al*. (2007) proposed 'A hotel recommendation system based on reviews: what do you attach importance to?'. They proposed a method to extract the preference of review contributors from a collection of reviews (Takuma *et al*., 2016).

Chen *et al*. (2011) senior member IEEE proposed the study "Travel recommendation by mining people attributes and travel group types from community-contributed photos, "they proposed to conduct personalized travel recommendation by further considering specific user profiles or attributes (e.g., gender, age, race) as well as travel group types (e.g., family, friends, couple). Instead of mining photo logs only, we exploit the automatically detected people attributes and travel group types in the photo contents (Chen *et al*., 2011).

Miraz *et al*. (2017) has proposed "collaborative web recommender framework for homestay programs where they used collaborative approach for recommending frameworks.

Kang *et al*. (2012) has proposed, "AWSR: active web service recommendation based on usage history" in this study they proposed a recommendation sytem where the history of the users are used for recommending services to the users.

Burke (2002) has proposed "hybrid recommender systems: survey and experi-ments, "user modeling and user-adapted interaction where a hybrid approach of recommender systems is used for the recommendation approach.

Meng *et al.* (2014) IEEE, has proposed "KASR: A keyword aware service recommendation method on mapreduce for big data applications" in this study a keyword aware approach is used as an algorithm for the recommendation of services to the users.

Jin *et al.* (2010) has proposed, "MyS-pace video recommendation with map-reduce on Qizmt, "here video recommendation approach is used for recommendation services.

Zhao and Shang (2010) has proposed, "User-based collaborative-filtering recommendation algorithms on hadoop where collaborative filtering approach is used for recommendation algorithm.

Gao *et al.* (2010) has proposed, "Solutions for problems of the existing e-Commerce recommendation system. "The improvements considered throughout this study mainly refer to the recommendation algorithm.

Baizal *et al.* (2016) has proposed, "ontology based recommendation involving consumer product reviews. "In this study, they elaborated the product reviews into the recommendation technique in addition to product features. Baizal *et al.* (2016) many customers are not familiar with the technical features of the product (e.g., Smartphone's, cars, cameras, notebooks, etc.,), so that, in the proposed framework, preference of customers or users is expressed in the form of functional requirements of products. So, there are many such works done in the field of recommendation systems and services.

However, a few more steps on the development ladder of recommendation systems are kept through this study.

MATERIALS AND METHODS

Materials used: The materials used for obtaining result are the review dataset of the users based on different hotels in different cities.

Methods used: The method used for obtaining result is the language modelling method which is given above in details.

Purpose of study: The main purpose of study on the service recommender system is to recommend the most appropriate recommendation list to the users. Recommendation systems play a very vital role in recommending the best appropriate services to the users. The field of recommendation system is very vast as there are many applications of the same. By recommending the best services to the users, the user can get the best deal in the field of his interest.

The best recommendation systems are those who provide the best out of best services to the users. So, for this reason in this study, to provide the best hotel recommendation list to the users, we are considering the users personalized requirements based on which the ratings and rankings of the hotels can be done.

Today the field of recommendation system has emerged a lot and due to the growing online information about customers and users this field is highly related with the big data environment. The big data environment poses many critical challenges on the recommendation systems, so, does it affects the services.

So on a positive end this study deals with the challenges faced on the recommendation system nowadays and tries to resolve them with great accuracy and efficiency. We aim at presenting the most appropriate services to the users according to their preferences. A keyword aware service recommendation system is presented in this study which process through the keywords of the requirements of the current user who is willing to use the recommendation system.

Another main objective of the study is to personalize the result of the project according to the user's preferences, so that, user gets the perfect result which he expects.

RESULTS AND DISCUSSION

In this study, the dataset contains the reviews of the distinct users from distinct hotels. There are multiple reviews for a single hotel. The data sets of reviews of different users are taken from www.tripadvisor.com. There are approximately 2, 00, 000 reviews of 10 different cities. The data set of reviews is in the form of date/short comment/long comment. This data set is, firstly, pre-processed. Pre-processing is done by applying, stemming, stopping and the lingo-refiner algorithms. Stemming is done using porter stemmer algorithm. After removing stop-words and doing stemming and lingo-refining, the words which are left are called the keywords (Fig. 2). These extracted keywords will be stored within the database along with the name of the hotel. The current user's preferences will be collected dynamically. The preferences of the current users will be taken as only keywords separated by commas. Each word separated by a comma will be considered as a keyword. Now, these keywords from the current user's preferences

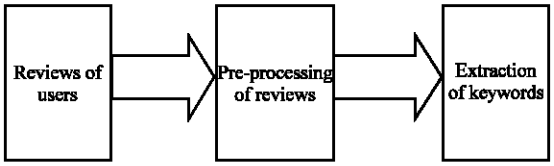

Fig. 2: Keyword extraction

and the keywords extracted from the reviews data set will be matched using the similarity computation algorithm. The algorithm which is used here is the language modeling algorithm.

Language modelling algorithm: Language modeling basically is probability distribution over sequence of words. Given a sequence as of length n, it assigns a probability P (w1, w2, ..., wn) to the whole sequence. Language modeling is used for many applications such as speech recognition, machine translation, part-of-speech tagging, parsing, handwriting recognition, information retrieval and other applications.

Language models are used as the query likelihood model in the field of information retrieval. Here, each document in a collection has a separate language model associated. Based on the probability of the query Q documents are ranked in the document's language model P (Q/Md). Commonly, the unigram language model is used for this purpose otherwise known as the bag of word's model.

Unigram models: A combination of several one-state finite automata used in information retrieval is called as unigram model. It splits the probabilities of different terms in a context:

$$P(t1\ t2\ t3) = P(t1)P(t2/t1)P(t3/t1t2)$$

To:

$$Puni(t1\ t2\ t3) = P(t1)P(t2)P(t3)$$

In this model, the probability to match (hit) each word depends on its own, so only one state finite automata as units is present. From the whole model, the sum of all the one-state hitting probabilities should be 1 (Table 1 and 2):

$$\sum \text{term index } P(\text{term}) = 1$$

The probability generated for a specific query is calculated as P(query) = $\prod$ term in query P(term). For different documents, we can build their own unigram models with different matching (hitting) probabilities of words in it. Example of unigram model in hotel recommendation system:

Table 1: Example of unigram model of a document

Terms	Probability in document
Frog	0.1
Toad	0.2
Said	0.05
Likes	0.05
That	0.3

Table 2: The probability of the whole sequence of words

Terms	Term frequency in document
Luxurious	0.2
Hotel	0.1
Resort	0.23
Swimming	0.52
Pool	0.52
Spa	0.4

p (luxurious hotel, resort, swimming pool, SPA) = 0.2*0.1*0.23*0.52*0.52*0.4 = 0.000497536

The terms of which probabilities are to be calculated are the keywords from the current user's preferences. Suppose the requirements of the current user are: luxurious hotel, resort, swimming pool and spa.

Then these requirements of the user will be considered as keywords. And the probability of these keywords will be calculated as the term frequency of these terms in a particular hotel (document):

Term Frequency (TF) = (No. of times term t appears in a document)/(Total No. of terms in the document)

The result obtained after multiplication is practically very small value, i.e., approximately (e to the power-17 or -19), so, we are taking the log of this value after multiplication. The value which we got after taking log can be considered as the score of hotel and depending on this score, the list which is to be recommended to the users can be ranked according to the hotels score and can be recommended to the users accordingly Algorithm 1.

Algorithm 1; Example result of language modelling algorithm:

Word match Sauna == spa
Word match Gym == gym
Hotel ==Admiral plaza
Final twp == 2.19675483677E-6
Log value == -5.657849282781

Use of domain ontology: A domain ontology is an ontology that describes all the relevant concepts in a single domain of interest. It is a knowledge framework of common concepts in a specific domain created by the domain experts. They capture valid for a particular type of domain. For example, electronic, medical, mechanic and digital domain. Domain ontology or domain-specific ontology models a specific domain or a part of the world. It represents the particular meanings of terms

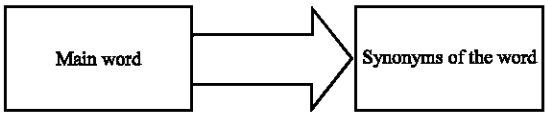

Fig. 3: Domain (Hotel) ontology (Example: SPA, health-facility, health-club, sauna, day-spa)

as they apply to that domain. Domain ontology is a collection of vocabularies and the specifications of the conceptualization of a given domain.

In our hotel recommendation system, we have created a dictionary which contains all hotel-related words (Fig. 3). The first word in all the rows are the base (root) word and the words besides the base words are the synonyms of that base word, i.e., if any user writes in his requirements the synonym of any word and that synonym is not present in the extracted keyword set of our reviews data set then, it will not say that the term does not exists, it will rather map that synonym to the root word and will fetch the hotel which contains the root word.

Example: If a user writes sauna instead of spa in his/her requirements (Sauna is the synonym of SPA). Then, our dictionary will map sauna with spa and say. Match word Sauna = SPA. And will fetch the hotel which contains spa keyword in it.

Priority of user: We can consider more specifically, the priorities of the user by taking the requirements of the user priority wise. Suppose if the user enters four requirements 1-4 and the requirements which are matched will display the hotels name. If all four requirements are matched, the hotel which matches the requirements will be at the top of the list. If three requirements are matched with some hotels, the hotel's name will be in the middle of the list to be recommended. And if two or one requirement is matched, the hotel will be least recommended or will be at the bottom of the list. Thus we propose a recommendation system which will try to recommend the most appropriate hotel according to the preferences of the user.

CONCLUSION

In this study, we have proposed a keyword aware hotel recommendation system. Here, keywords are used to indicate user's preferences and a language modeling algorithm is adopted to generate the most appropriate recommendations. More specifically, a dictionary containing hotel-related words is used to help recommend more appropriate services. The current user gives his/her preferences in the requirements and those requirements are matched with the extracted keywords of the reviews data sets. The method aims at presenting a personalized service recommendation list and recommending the more appropriate services to the user.

RECOMMENDATIONS

In the future researcher, we will do further research in how to distinguish the positive and negative preferences of the users from their reviews to make the predictions more accurate.

REFERENCES

Adomavicius, G. and Y. Kwon, 2007. New recommendation techniques for multicriteria rating systems. IEEE. Intell. Syst., 22: 48-55.

Asawarangsee, T. and S. Maneeroj, 2016. A novel aggregation technique for multi-criteria recommendation. Proceedings of the 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), July 13-15, 2016, IEEE, Khon Kaen, Thailand, ISBN:978-1-5090-2034-8, pp: 1-6.

Baizal, Z.K.A., A. Iskandar and E. Nasution, 2016. Ontology-based recommendation involving consumer product reviews. Proceedings of the 4th International Conference on Information and Communication Technology (ICoICT'16), May 25-27, 2016, IEEE, Bandung, Indonesia, ISBN:978-1-4673-9880-0, pp: 1-6.

Bjelica, M., 2010. Towards TV recommender system: Experiments with user modeling. IEEE. Trans. Consum. Electron., Vol. 56

Burke, R., 2002. Hybrid recommender systems: Survey and experiments. User Model. User-Adapted Interact., 12: 331-370.

Chen, Y.Y., A.J. Cheng and W.H. Hou, 2011. Travel recommendation by mining people attributes and travel group types from community-contributed photos. Proceedings of the 19th ACM International Conference on Multimedia, November 28-December 01, 2011, ACM, Scottsdale, Arizona, pp: 1-2.

Gao, Z., Z. Li and K. Niu, 2016. Solutions for problems of existing E-Commerce recommendation system. Proceedings of the 2016 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW'16), May 27-29, 2016, IEEE, Nantou, Taiwan, ISBN:978-1-5090-2074-4, pp: 1-2.

Gobin, B.A. and R.K. Subramanian, 2007. Knowledge modelling for hotel recommendation system. Intl. J. Soc. Behav. Educ. Econ. Bus. Ind. Eng., Vol. 1,

Jin, Y., M. Hu, H. Singh, D. Rule and M. Berlyant *et al*., 2010. MySpace video recommendation with map-reduce on Qizmt. Proceedings of the 2010 IEEE 4th International Conference on Semantic Computing (ICSC'10), September 22-24, 2010, IEEE, Pittsburgh, Pennsylvania, ISBN:978-1-4244-7912-2, pp: 126-133.

Kang, G., J. Liu, M. Tang, X. Liu and B. Cao *et al*., 2012. AWSR: Active web service recommendation based on usage history. Proceedings of the 2012 IEEE 19th International Conference on Web Services (ICWS'12), June 24-29, 2012, IEEE, Honolulu, Hawaii, ISBN:978-1-4673-2131-0, pp: 186-193.

Linden, G., B. Smith and J. York, 2003. Amazon. com recommendations: Item-to-item collaborative filtering. IEEE. Internet Comput., 7: 76-80.

Meng, S., W. Dou, X. Zhang and J. Chen, 2014. KASR: A keyword-aware service recommendation method on mapreduce for big data applications. IEEE. Trans. Parallel Distrib. Syst., 25: 3221-3231.

Miraz, M.H. R. Ramli and K.R.K. Mahamud, 2017. Collaborative web recommender framework for homestay programs. J. Eng. Appl. Sci., 12: 1575-1581.

Sanchez, F., M. Alduan, F. Alvarez, J.M. Menendez and O. Baez, 2012. Recommender system for sport videos based on user audiovisual consumption. IEEE. Trans. Multimedia, 14: 1546-1557.

Takuma, K., J. Yamamoto, S. Kamei and S. Fujita, 2016. A hotel recommendation system based on reviews: What do you attach importance to?. Proceedings of the 4th International Symposium on Computing and Networking (CANDAR'16), November 22-25, 2016, IEEE, Hiroshima, Japan, ISBN:978-1-5090-2656-2, pp: 710-712.

Zhao, Z.D. and M.S. Shang, 2010. User-based collaborative-filtering recommendation algorithms on hadoop. Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (WKDD'10), January 9-10, 2010, IEEE, Phuket, Thailand, ISBN:978-1-4244-5397-9, pp: 478-481.