A Personalized Recommendation System Using Collaborative Filtering With Feature Based Sentiment Analysis

Nyein Ei Ei Kyaw, Thinn Thinn Wai, Thiri Haymar Kyaw
University of Information Technology, Yangon,Myanmar
nyeineieikyaw@uit.edu.mm, thinnthinnwai@uit.edu.mm, thirihaymarkyaw@uit.edu.mm

Abstract

Recommendation systems help users to deal with the information overload problem by producing personalized content according to their interests. For presenting the personalized recommendation according to the new user's demand is big challenge. Beyond the traditional recommender strategies, there is a growing effort to incorporate users' reviews into the recommendation process, since they provide a rich set of information regarding both items' features and users' preferences. This proposal proposes a recommender system that uses users' reviews and preference of new users to meet the individual interest.

Keywords- Recommender system, Preferences, Review

1. Introduction

Today, people currently live in an era of information. They are surrounded by a excess of data in the form of reviews, blogs, papers and comments on various websites. With the success of the Web 2.0, more and more companies capture large-scale information about their customers, providers, and operations. In recent years Recommender systems are software applications that attempt to reduce information overload by recommending items of interest to end users based on their preferences. With the growing number of alternative services, effectively recommending services that users preferred have become an important research issue. Service recommender systems have been shown as valuable tools to help users deal with services overload and provide appropriate recommendations to them. Examples of such practical applications include CDs, books, web pages and various other products now use recommender systems. Over the last decade, there has been much research done both in industry and academe on developing new approaches for service recommender systems.

2. Related Works

Several papers have addressed the problems to meet the personalized requirement of user in various ways.

Pallavi R. Desai, B. A. Tidke proposed a system that present personalized recommendation list and recommend the most appropriate items to the user by using weights of keywords are used to indicate user' preferences and a user-based collaborative filtering algorithm is adopted with Opennlp to generate appropriate recommendations [1]. Khushboo R. Shrote, Prof. A.V. Deorankar proposed a system in which feedback analysis is done using sentiment analysis to recommend services. Keywords are used to indicate what the users prefer [2]. Susan Thomas, Jayalekshmi S proposed a system in which sentimental analysis on the reviews is done using Naïve Bayes, a machine learning technique to distinguish between the positive and negative reviews. It also use MonoDB database to store the review detail [3].Shakhy.P.S1, Swapna.H2 proposed a recommendation system which considers not only user reviews but also the temporal information about location of the services. It use Apache Mahout learning library and MongoDB to store reviews [4]. Dr. Kogilavani Shanmugavadivel and their colleagues proposed a system deals with the implementation of personalized rating to the services for hotel reservation system and booking of cars. This system performed opinion mining on the review at the sentence level using Bayes theorem and negation rule algorithm [5].

All the above papers based on the previous users' reviews and new user preferences are considered as keyword. Sentiment analysis is performed on the reviews by using machine learning algorithms and then similarly of previous users' keyword set and active user keyword set are computed and finally recommend the top k services to the new user. The candidate service set that system provide and similar terms associated with candidate service are manually specified.

Therefore, this proposal intends to construct the domain ontology concerned with terms and related terms of the domain and perform sentiment analysis on the review and recommend the top k services to the user in order to meet the need of user more accurately. This proposal consists of the following parts: (1) A user-based collaborative filtering recommendation system, (2) domain ontology is constructed for features identification of the domain (3) Sentiment analysis on the reviews is done using Lexicon based approach.

3. Background Theory

Personalized recommender systems help users to find a wide variety of products online, and assist users in making decisions. At such highly rated Internet sites as Amazon.com, YouTube, Netflix, Yahoo, TripAdvisor, Last.fm, and IMDb, recommender systems play a very important role. Users could take advantage of these recommender systems to find a variety of products, videos, books, and news that they like from the massive available item set [8]

3.1 Recommendation System

Recommendation System, a sub-class of information filtering system, helps in predicting top-N preferred items for a user. Recommendation system can be classified into content based approach, collaborative filtering approach and hybrid approach. Content based recommendation systems will recommend items based on the description of the items and profile of the user. Collaborative filtering recommendation system will recommend items based on the similarity between the users who have rated the same item before. This methods build a model using information about past purchases or ratings provided by users. This model may be used for prediction of preference rating for a given item. Hybrid approach is a combination of content based and collaborative filtering approaches.

3.2 Collaborative Filtering (CF)

Collaborative filtering methods analyze large amount of information about preferences of users and predict preferences of similar users for recommending items. Collaborative filtering (CF) methods are of two types: item based and user based collaborative filtering. Item based CF, recommend items based on the similarity between the items rated by the same user in the past. User based CF, recommend items based on the similarity between the users who have rated the same items.

3.3 Similarity Computation

Similarity computation between items or users is a critical step in collaborative filtering algorithms. There are many different methods to compute similarity or weight between users or items such as Pearson correlation, Vector Cosine-Based Similarity, Euclidean distance, Minkowski distance, Cosine similarity.

3.4 Sentiment Analysis

Opinion mining (Sentiment Analysis) is a Natural Language Processing (NLP) and Information Extraction (IE) task that aims to obtain the feelings of the writer expressed as positive, negative or neutral opinions by analyzing large number of documents.

Sentiment Classification techniques can be divided into machine learning approach, lexicon based approach and hybrid approach. Machine learning approach is based on machine learning algorithms. Lexicon based approach is based on sentiment lexicon and hybrid approach combines both approaches. Three level of sentiment analysis are:

1. Document Level: The whole file contains group opinion. The file verifies whether it conveys the positive or negative sentiment.
2. Sentence Level: Verifies that whether a sentence conveys positive, negative or neutral meaning.
3. Entity and Aspect Level: Product features are defined as product attributes or components. Analysis of such features for identifying sentiment of the document is called as feature based sentiment analysis. In this approach positive or negative opinion is identified from the already extracted features.

3.5 Ontology

Ontology is a formal representation of the knowledge by a set of concepts within a domain and the relationships between those concepts. It is used to represent the properties of that domain and may be used to describe the domain. Domain ontology is one kind of ontology which is used to represent the knowledge for a particular type of application domain (e.g., a product domain).

Domain-Ontology structure mainly consists of several attributes of ontology concept. Concept is the abstract of concrete objects. The attributes of concept are used to describe the characteristics of various aspects of concept. For example, film is a domain concept, director, actor, and released age are attributes of the concept [6].

4. Overview of the system

In the proposed system, the reviews of specific domain are firstly collected. Then preprocessing step is carried out on the review. In this stage stop words and HTML tags are removed from the reviews. Then the words are converted to the root forms of the word. That is words such as "computing", "computation" and "computes" have the root form as "compute". This process of getting the root form of the words is called stemming. Porter Stemmer algorithm is used for stemming.

After that, domain ontology is constructed to collect the terms and related similar terms of the domain. It contains the candidate services and reference work of the candidate service set which contains terms with similar meanings of the specific domain. The system then performs the opinion mining step to assign the polarity of each feature from the domain ontology based on SentiWordnet. Similarity computation of previous user and active user preference are performed on the keyword set of both users using Cosine similarity measures. Finally, personalized rating is calculated and top k recommendation lists are presented to the user.
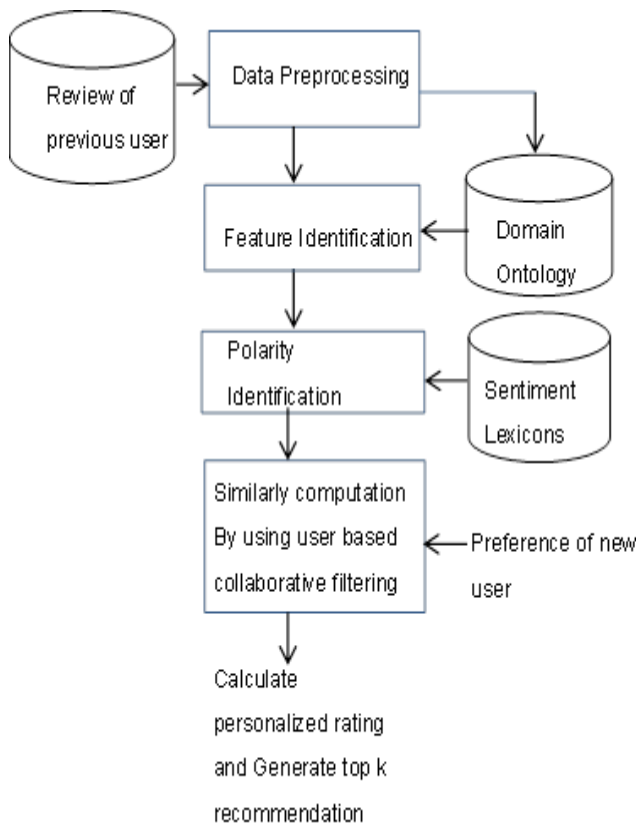

Figure1. Proposed system architecture

5. Conclusion

Most existing service recommender systems are only based on a single numerical rating to represent a service's utility as a whole. In fact, evaluating a service through multiple criteria and taking into account of user review can help to make more effective recommendations for the users. Therefore, the proposed system will be implemented the personalized recommendation system by constructing domain ontology and sentiment analysis on the review to meet the individual user interest.

6. References

[1] Pallavi R. Desai, B. A. Tidke , "A Survey on Smart Service Recommendation System by Applying Map Reduce Techniques", International Journal of Science and Research (IJSR) 2014

[2] Khushboo R. Shrote, Prof. A.V. Deorankar, "Sentiment Analysis Based Feedback Analysed Service Recommendation method For Big Data Applications", International Journal of Scientific & Engineering Research 2016

[3] Susan Thomas, Jayalekshmi S., "Recommendation System with Sentimental Analysis using Keyword Search", international journal for advance research in engineering and technology 2015

[4] Shakhy.P.S1, Swapna.H2,"Improved Keyword Aware Service Recommendation System for Big Data Applications", International Journal of Innovative Research in Computer and Communication Engineering 2015

[5] Dr. Kogilavani Shanmugavadivel, Dr. Thangarajan Ramasamy , Dr. Malliga Subramanian," Semantic Ranking Based Service Recommendation System using MapReduce on Big Datasets" , International Journal of Advances in Computer and Electronics Engineering 2017

[6] Cheng Xiao, Dequan Zheng, Yuhang Yang, Automatic Domain-Ontology Structure and Example Acquisition from Semi-Structured Texts Sixth International Conference on Fuzzy Systems and Knowledge Discovery 2009

[7] Saikat Bagchi, "Performance and Quality Assessment of Similarity Measures in Collaborative Filtering Using Mahout", 2nd International Symposium on Big Data and Cloud Computing

[8] F. Ricci, L. Rokach, B. Shapira, P.B. Kantor (Eds.), Recommender Systems Handbook, Springer, 2011, http://www.springerlink.com/content/978-0-387-85819-7.