



A Review on Content Based Recommendation System

Deepa Mandave, Govind S. Pole

M. E Student, Department of Computer Engineering, MES College of Engineering, Pune, India

Assistant Professor, Department of Computer Engineering, MES College of Engineering, Pune, India

ABSTRACT: Multiple Options available on the internet, there is requirement to sift, rank and professionally provide relevant material. Many Internet users have the problem of information overload. To reduce this problem, there is requirement of one solution. Recommendation systems resolve this problem by examining through huge capacity of dynamically produced data to deliver scholars with personalized content and services. Most of research has been done on the collaborative and hybrid recommendation system, content based system has not been that extensively researched. This paper presents a literature review of some papers on content based recommendation system. Implementation of a content based recommendation contains comparing the features of a user profile, with the characteristics of a content item, to predict to the user novel relevant items. This research delivers information about developments in content-based recommendation systems and offers experts with understanding and future scope on content-based recommendation systems.

KEYWORDS: Collaborative and hybrid recommendation system, content-based recommendation systems, personalized, Recommendation systems.

I. INTRODUCTION

Now a days, E-commerce technology is very famous for information explosion. Most studies annoyed to develop the autonomous system which identifies the user's desires. Most popular tool that helps users to recommend according to their interests is Recommendation System (RS).

Recommendation systems have become significant research area since the arrival of the first research paper on collaborative filtering in the mid-1990s. Academic research on recommendation systems has increased over the past 15 years, there are shortages in the comprehensive literature review and grouping of that research [1]. Recommendation systems help consumers to find what they really want. So this meets the desires of consumers in a short time. It helps consumers to find information, products, or by gathering and exploring suggestions from other users actions [3]. There are plenty of choices and a slight time to explore them all. Recommendation systems service people make choices in these composite information spaces. Recommendation systems are a category of information filtering that represents lists of items (films, songs, books, videos, images, products, web pages...) which are similar of user interest. Most popular recommendation systems in all over the world are Amazon, Last.fm, Ulike, iLike, Netflix, Pandora etc. Simply they match user interest learned from his/her profile with some reference features and guess the rating that the user would give. Those features may be from the item information (content-based method) or the user's social profile (collaborative filtering method) [4].

The main objective of recommendation systems is to help users to deal with the information burden problem by delivering personalized recommendations, content and service. Recommendation systems are progressively being used in E-commerce for recommending books, music, movies, TV shows or different types of objects [6]. Recommendation systems are essential tools that overcome the information overload by examining through the large set of data and recommending information relevant to the user. In general, every recommendation system follows a particular process [1] in order to create product recommendations, see Figure 1. Collaborative Filtering, Content based method and Hybrid methods are the widespread three methods to developing recommendation systems. In Collaborative Filtering (CF) method, thoughts from users in the form of rankings on several items are collected. The recommendations generated are based only on the thoughts of users similar to the active user. Active user refers to recommendation searcher for whom recommendations are produced. Content based approach proposes items that are matches to the ones the active user has shown a favorite for in the past rather than on the favorites of other users[5]. CBF methods are

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

developed for information retrieval and information filtering study. In the CBF method, each user can work independently and will be recommended the most closely information of the objects according to their request[3]. This paper, focuses on fully and systematic learning of content based recommendation approach. The main goal is :

- 1) An outline of this approach is to concentrate on the effective techniques and application fields that realized these techniques;
- 2) Developments and guidelines for future work is useful for next level of content-based recommendation approach.

Approximately all the study has been done on collaborative or hybrid recommender system but very less work is done on content based recommendation system [12]. This paper delivers an outline of content-based recommendation approach, by keeping the objective to show multiple perspectives in design and practical scenario. Initial section related to the paper represents the fundamental aspects and terms of content based recommendation approach, a high level structure, and key benefits and disadvantages. Following section in the paper presents a survey of current systems used in numerous application field, by comprehensively describing both traditional and innovative methods for presenting objects and scholar profiles. The most broadly accepted methods for learning scholar profiles are also explored. The final section of the paper argues developments and future work.

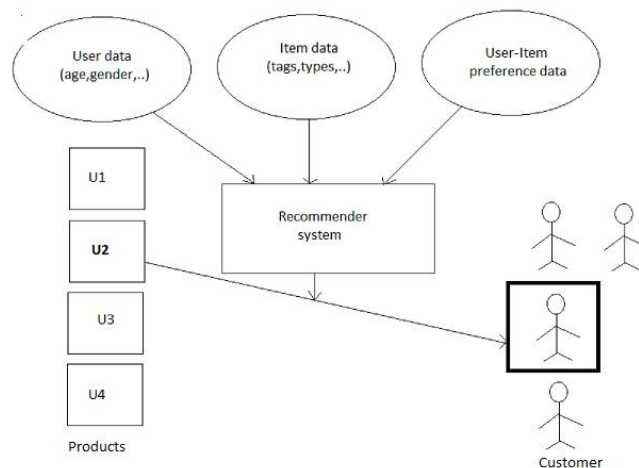


Figure1. Recommendation Process

II. CATEGORIZED RELATED WORK

Content based filtering analyzes the insight of data objects to forecast its relevance which depends upon scholar's profile. Many computer science subjects such as Information Retrieval and Artificial Intelligence include study of content based recommendation approach. Recommendation technologies are actually used Information Retrieval (IR) techniques which includes users findings are as an information pursuing process for recommendations. Items which will be recommended can be diverse based on number and kinds of features used in recommending process.

In case of Artificial Intelligence, recommendation job consider as a learning problem which analyzes earlier information about scholars. Scholar profiles consist of user-specified keywords, guidelines or rules and affect the lengthy favorites of the scholar. This commonly includes the application of Machine Learning (ML) methods.

A. VECTOR SPACE MODEL (KEYWORD-BASED MODEL)

Maximum content based recommendation approach use comparatively basic retrieval method, such as keyword equivalent or the Vector Space Model (VSM) which includes classical weighting TF-IDF. Generally, each document contain terms, so it is having vector of term weights, that every weight specifies the degree of relation between term and the document. Let $D = \{d_1, d_2, \dots, d_n\}$ presents a list of documents or repository, and $T = \{t_1, t_2, \dots, t_n\}$ represents the dictionary, that means the list of keywords in the body of document. T is achieved by delivering some traditional operations of natural language like stopwords removal, tokenization, stemming etc.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

$$\text{TF-IDF}(t_k, d_j) = TF(t_k, d_j) \log \frac{N}{n_k}$$

Where N represents the total number of documents in group, and n_k represents the number of documents in the group which contain the term t_k at least once. To define the two documents nearness or closeness, a match measure is needed. Various similarity metrics have been achieved to demonstrate the closeness of two sets; in those metrics, one of the extensively used is cosine similarity:

$$i(d_i, d_j) = \frac{\sum_k w_{ki} \cdot w_{kj}}{\sqrt{\sum_k w_{ki}^2} \cdot \sqrt{\sum_k w_{kj}^2}}$$

In content based recommendation approach, both scholar profiles and items are denoted as weighted term vectors using VSM. Cosine similarity is used to compute user's favorites in a certain item required for prediction.

B. SURVEY OF KEYWORD BASED SYSTEMS

Many keyword based recommendation approach have been found in many areas of applications such as web pages, news, movies, e-commerce, music etc. Every domain represents diverse problems that need diverse solutions.

In food recommendation system, Ipek Tatli proposed content-based food recommendation system. It contains three distinct parts; food profiling, user profiling and recommending foods according to the earlier feedback of the users. Food domain can be realized as a set of features where each feature is essentially an ingredient of that food. Hence, both item and user profiles are stored as vectors of features. It attempts to discover best matches between the user profile and the food profiles. TF-IDF distance measure used to calculate the match of two items. But it gives poor recommendation that is evaluation logic may also be leading to some failures [16].

Keyword Based TV Program Recommendation by Christian Wartena, et al. has explored keyword based ranking prediction. An appropriate level of description of object established by Keywords since mined automatically from one or more texts or keywords can be specified by individuals. They have presented that keyword based ranking recommendation provide very suitable results for some datasets, comparable to current development in collaborative filtering approach. Dataset nature and relevance of domain of item are reasons for the item appreciation assumed by authors [18].

Social tags increases a significant attention or interest of many researchers for recommendation. In the area of social tagging system, Iván Cantador et al. proposed several content based recommendation representations that uses scholar and item profiles represented as a weighted set of social tags. The planned methods are implementations of the Okapi BM25 and Vector Space ranking models. Actually, recommendation models constructed using whole tag-based training items by this system. It also contains user's profiles based on tags made by tags explaining the training objects. We calculated the recommendation system with the tag-based profiles of the test items. In the evaluation, they calculated numerous metrics, and implemented a 5-fold cross validation method. This is effective and efficient but the time for training is a serious issue in this scheme [17].

Ei Hlaing proposed car recommendation system using the relevant words extraction method from users email. Rule-based method used by this work the in Compiling procedure. The most appropriate for relevant words extraction is Context-free grammar technique. This work performs recommendation efficiently by using proposed key extraction algorithm with Content-based approach and Jaccard Coefficient. By delivering significant car information, this system will benefit the users who need to purchase the car. This study posses higher precision with less recall compared to machine learning method [3].

Text classifiers mostly contain content based approaches. Text classifiers are prepared by training sets having documents containing positive or negative illustrations of user favorites. Hence, precise recommendations are



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

generated when large numbers of examples are available in the training sets, which promise trustworthy “syntactic” indication of user favorites. “Lack of intelligence” is the issue with keyword based method. When additional innovative features are needed, keyword based methods display their restrictions. For example, when the user, prefer “French impressionism”, keyword based methods will just identify documents containing the keywords “French” and “impressionism”. More innovative representation approaches are required to prepare content based recommendation approach using “semantic intelligence”, that permits going away from the syntactic indication of user likeness delivered by words. In the succeeding parts, to test probable means to fill facts using ontologies and encyclopedic information sources.

C. SEMANTIC ANALYSIS USING ONTOLOGIES

Semantic exploration permits learning additional correct profiles which hold links to thoughts stated in exterior information sources. Recommendation system with linguistic and cultural background information that indicates the capability of interpreting natural language documents and focusing on their content is the challenge and main motivation for this system. In this part, a survey of the key policies implemented to announce certain semantics to the recommendation practice is denoted. The explanation of those approaches is accepted out by considering some conditions:

- The kind of information source used (e.g. ontology, lexicon, etc.);
- Methods implemented for the representation and annotation of the objects;
- Insight of scholar profile;
- Item-profile comparing scheme.

In the field of art, Yiwen Wang et al. verified “Art Recommender”, CHIP demonstrator, which generated recommendation by exploring artworks features (e.g. material, creator and subject) and semantically interrelated concepts with active users. To indicate how personalization strategies and Semantic Web can be used to improve access to museums of digital gatherings is the key objective of the Cultural Heritage Information Personalization (CHIP) project [7].

ANTSREC-A Semantic Recommendation System using Ant Colony (Meta-Heuristic) algorithm in E-Commerce by Mojtaba Salehi, et al. proposed a semantic recommendation system depends on Ant Colony algorithm which used ontology dependences for enhancement of E-commerce. This work includes reinforcement learning, stochastic, heuristic within Ant Colony algorithm and semantic dependence in ontology features. The proposed work is capable of recommending bundled, complement and similar items. These features may overcome difficulties such as scarcity of information, cold start and scalability[1].

With the rapid increasing of learning objects (LOs) in a variety of media formats, it requires to assist learners in finding the appropriate LOs which will be needed for their learning. Salam Fraihat et al. have proposed a framework of a semantic recommendation system for e-learning in which it will help learners to discover and select the relevant LOs to their area of interest. The proposed framework utilizes the intra and extra semantic relationships between LOs and the learner’s needs to provide personalized recommendations for learners. The semantic recommendation algorithm is based on the extension of the query keywords by using the semantic relations, concepts and reasoning means in the domain ontology. The proposed system can be used to reduce the time and effort involved in finding suitable LOs, and thus, improves the quality of learning [6].

All researches which integrated any domain specific information, linguistic or both in content based approaches displayed improved and more precise outcomes compared to classical content based approaches. This work boosts investigators to plan new filtering approaches that formalize, contextualize user favorites by manipulating exterior information sources as ontologies or thesauri.

III. FUNDAMENTALS OF CONTENT-BASED RECOMMENDATION SYSTEMS

In Content based approach, recommendation of objects is related to qualities or insight of objects rather than rankings of users as shown in Figure 2. Content-based filtering method highlights more on the characteristics of items to produce recommendations. Content based approach is the best in predicting documents like web pages, publications, articles

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

and news etc. In content based method, prediction is related to scholar profiles with attributes mined from the insight of the objects, the scholar has intended in the historical. Positively rated objects or items that got high rank are recommended to the user [9].

Systems realizing a content based recommendation method evaluate a collection of articles and attributes of objects ranked by a scholar in the past, and construct a design or profile of scholar interests using attributes of objects ranked by scholar. The user profile is a picture of user likeness in structured manner, implemented to predict novel relevant objects. The recommendation method fundamentally consists comparing the features of the scholar profile with the characteristics of a content element[8]. Content-based recommendation systems (e.g. Last.fm6 and Amazon7) explore item features to find items which are similar to favorites of the user. Content based recommendation approach sift large repositories of objects (e.g. TV assets, articles, news, web pages, books, music tracks) with considering objects previously ranked by a scholar and constructing a analyze model of scholar favorites, stated as scholar profile, built upon the attributes of the objects ranked by the scholar. Then the scholar profile is analyzed for recommending novel possibly relevant objects. Textual attributes are commonly used to describe objects and scholar profiles by CBRS; therefore this produces usual problems of uncertainty in natural language. An expressive recommendation generated by CBF explores diverse kinds of models to identify match between articles. Vector Space Model like Term Frequency Inverse Document Frequency (TF/IDF) or Probabilistic models like Naive Bayes Classifier, Neural Networks or Decision Trees could be used to construct the relationship between different documents within a repository. To learn the basic model, statistical study or machine learning methods are used for making recommendation. Different scholar profiles does not affect recommendations generated by Content based method. CBF method has the potential to adjust its recommendations with changes in user profile within a short time. The main drawback in this method is the requirement of deep knowledge with exploration of the features of the object in the scholar profile [9].

Recent growth in the field of CBRS through semantic based techniques with numerous open sources interest in this field is evolved. New research study has presented semantic based methods that move a keyword based illustration of objects and scholar profiles to a conception based[3].

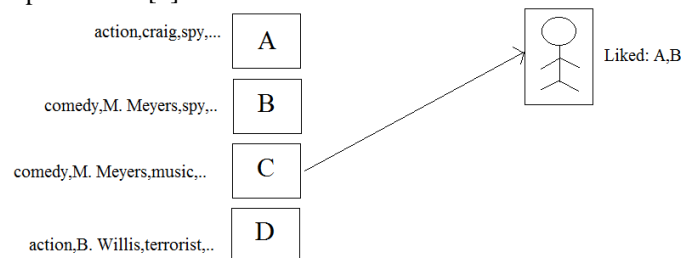


Figure 2 Content-based approaches

A BENEFITS AND DISADVANTAGES OF CONTENT BASED APPROACH

Content based recommendation approach has numerous advantages as related to the collaborative one:

- **USER INDEPENDENCE:** Content based recommendation system analyze uniquely rankings delivered by the active scholar to construct their profile. Else, collaborative filtering approaches require rankings from different users to identify “closest neighbors” of the current scholar.
- **TRANSPARENCY:** Content based approach displayed explicitly set of content features or a characteristic that matches with recommended items for active user. This creates trust in recommendation with evidence such as those features. Instead, in collaborative systems, like black boxes recommendation is displayed by collecting same choices of different strange users.
- **NEW ITEM:** In collaborative recommendation system, new item can be recommended when it is ranked by any user, otherwise it cannot be recommended. Content based recommendation does not have this cold start problem. It has capability of recommending objects not still ranked by any user.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

Content-based systems have some limitations:

- **RESTRICTED CONTENT ANALYSIS** – In the Content-based systems, recommended objects are limiting in number and kind of attributes those are related, either automatically or physically. Domain knowledge is often desirable, for example, movie recommendations that contain scheme necessities to recognize the artists, producers and field ontologies are sometimes required.

SERENDIPITY – Content based recommendation have no integral method for finding somewhat unpredicted or unique taste of user. This system recommend objects which having large score when matched with the scholar profile; for example, scholar has only ranked movies produced by Stanley in the past, he/she will get recommendations only those movies.

- **NOVEL USER:** Sufficient features must be gathered before content based recommendation approach can actually know user likings and deliver correct recommendations. Hence, if there are few rankings available for a novel user, the approach can not to deliver assured recommendations.

Some approaches for handling the stated problems will be represented in the next section.

IV. TRENDS AND FUTURE RESEARCH

Content based filtering method is a domain-dependent and it highlights largely on the exploration of the features of items to produce predictions. Content based filtering approach is the utmost successful method in predicting documents such as news, web pages and publications. CBF method has the capability to change its recommendations with changes occurs in user profile within a very short period of time. CBF uses diverse types of prototypes to discover match between articles or documents for producing meaningful recommendations like keyword based model (Vector Space Model) for example, TF/IDF (Term Frequency Inverse Document Frequency). “Lack of intelligence” is the problem with this method. When extra innovative features are necessary, keyword based methods display restrictions. More advanced techniques are required to prepare content-based recommendation system with “semantic intelligence”.

With semantic analysis, it is a possible learning accurate profile that includes links to aspects mentioned in exterior information sources. The challenge for this method is the key inspiration. The challenge is to provide a recommendation system which has the linguistic background information and cultural information that describes the capability of analyzing natural language text documents with arguing on the content. In building the application field, ontologies show the major character that being deployed for the presentation of the aspects and for semantic descriptions of the objects and associations recognized in the domain. There is a method that is capable to deliver natural language documents with semantic representation at fine-grained level in a large quantity of natural thoughts obtained from Wikipedia is called ESA (Explicit Semantic Analysis).

The current concerns in text representation using world knowledge sources for example, Wikipedia, have been focused, though that has not still deployed in the environment of learning scholar/user profiles.

V. CONCLUSION

We reviewed the area of content based recommendation approach, by presenting survey of the utmost significant concepts describing that type of methods. They share general method for representing objects to be recommended, although there is collection of recommendation systems in diverse fields. We explored the main content recommendation systems, by underlining the causes for that a more critical “semantic analysis” for content is required to go away from the syntactic indication of user favorites delivered by keywords. A review of recommendation methods which used semantic approaches is conducted, by delivering the linguistic knowledge as the primary role. The final section in the paper is indicated key developments and study of content based recommendation approach for the next researcher. We concluded that all research that is related with domain specific information or linguistic or both in content based methods provided more precise and better outcomes compared to classical content based methods. This inspires investigators to propose new filtering approaches which built and contextualize user favorites by using exterior information sources like ontologies or thesauri.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

REFERENCES

1. Mojtaba Salehi, Abdolhossein Fathi and Fardin Abdali-Mohammadi , “ANTSREC: A Semantic Recommender System Based on Ant Colony Meta-Heuristic in Electronic Commerce”, *International Journal of Advanced Science and Technology*, Vol. 56, July, 2013.
2. Michael J. Pazzani and Daniel Billsus , “Content-based Recommendation Systems”.
3. Pasquale Lops , “Semantics-aware Content-based Recommender Systems”.
4. Salam Fraihat, Qusai Shambour, “A Framework of Semantic Recommender System for e-Learning”.
5. Punam Bedi, Ravish Sharma, “Trust based recommender system using ant colony for trust computation”, *Expert Systems with Applications* 39 (2012) 1183–1190.
6. Yiwen Wang, Natalia Stash, Lora Aroyo, Laura Hollink, and Guus Schreiber “Using Semantic Relations for Content-based Recommender Systems in Cultural Heritage”.
7. Yiwen Wang Natalia Stash Lora Aroyo Laura Hollink Guus Schreiber , “Semantic Relations in Contentbased Recommender Systems”.
8. Pasquale Lops, Marco de Gemmis and Giovanni Semeraro , “Content-based Recommender Systems: State of the Art and Trends”.
9. F.O. Isinkaye , Y.O. Folajimi , B.A. Ojokoh, “REVIEW Recommendation systems: Principles, methods and evaluation”, *Egyptian Informatics Journal* (2015) 16, 261–273.
10. ” Deuk Hee Park, Hyea Kyeong Kim, Young Choi, Jae Kyeong Kim, “A Review and Classification of Recommender Systems Research”, 2011 *International Conference on Social Science and Humanity*.
11. RVVSV Prasad and V Valli Kumari, , “A CATEGORICAL REVIEW OF RECOMMENDER SYSTEMS”, *International Journal of Distributed and Parallel Systems (IJDPSS)* Vol.3, No.5, September 2012.
12. Parul, Kavita khanna , “Literature Survey: Recommender Systems “, *IJETMAS*, March 2015, Volume 3 Special Issue, ISSN 2349-4476
13. Amir Hossein Nabizadeh Rafsanjani, Naomie Salim, Atae Rezaei Aghdam, Karamollah Bagheri Fard , “Recommendation Systems: a review” , *International Journal of Computational Engineering Research*||Vol, 03||Issue, 5||
14. Daniar Asanov , “Algorithms and Methods in Recommender Systems”.
15. “Research-Paper Recommender Systems: A Literature Survey” ,Joeran Beel, Bela Gipp, Stefan Langer, and Corinna Breiting.
16. Ipek TATLI , “Food Recommendation System”, 1395557, June2009.
17. Iván Cantador, Alejandro Bellogín, David Vallet , “Content-based Recommendation in Social Tagging Systems”.
18. Christian Wartena, Wout Slakhorst, Martin Wibbels, Zeno Gantner, Christoph Freudenthaler, Chris Newell, Lars Schmidt-Thieme , “Keyword-Based TV Program Recommendation”.
19. Sweta Sarkar, Pritha Ghosh , “A Hybrid Collaborative Filtering Approach for Recommendations “, *IJETAE*, ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 5, Issue 2, February 2015.
20. LIWEI LIU, FREDDY LECUE, INIKOLAY MEHANDJIEV , “Semantic Content-based Recommendation of Software Services using Context”