

## **A Review Paper on Recommender Systems To Optimize Search Space and Sparsity In E-Commerce Environment**

**Mr. T. Subha Mastan Rao<sup>1</sup>, Ms.A.Prasanna<sup>2</sup>, Mr.R.Vikram Singh<sup>3</sup>, Mr.B.Rahul<sup>4</sup>, and Mr.K.Rahul<sup>5</sup>**

*Asst Professor<sup>1</sup>,*

*cse,Department,KLUniversity,vaddeswaram,mail:mastan@kluniversity.in*

*Student<sup>2,3,4,5</sup>, Department, KLUniversity, vaddeswaram,*

*ambatiprasanna64@gmail.com<sup>2</sup>*

*vicky45263@gmail.com<sup>3</sup>, rahulchowdary888@gmail.com<sup>4</sup>,*

*kollirahul111@gmail.com<sup>5</sup>*

### **Abstract**

Recommender systems assist users in retrieving the interesting items by automatically acquiring user preferences from the historical data of the user and matching items with his preferences. In the last decade, there is exponential increase in the information available in the internet. For every product there are various choices available to the user. So the recommender systems gain a lot of interest these days. In this paper we are going to discuss about various types of recommendation techniques available like content based technique, collaborative technique, hybrid, demographic, knowledge based and utility based techniques. Even though there are advantages in each technique but there are a few shortcomings too. Each technique has a different approach. Content based technique is completely dependent on user preferences but collaborative approach will consider the preferences of similar users. Hybrid technique is a combination of both collaborative and content based technique. Demographic data of the user is considered in demographic based technique where as knowledge based approach requires relation of user preferences with his need. The main goal of this paper to describe features of different recommendation techniques available and compare them. Limitations observed in one technique can be overcome in other techniques. Although a hybrid recommendation technique gives effective results we have to choose the technique depending on the requirement.

**Keywords:** Recommender systems, Collaborative technique, Content based technique, utility based, Demographic recommender systems, knowledge based recommender systems.

## Introduction

The amount of information present in internet is increasing day by day. Retrieving of relevant information from available data is becoming a difficult task because we have to compare all the related data and have to get best results out of it. . With the extensive development of web applications, there is also a rapid increase in online services that are becoming popular, focusing to provide different types of information based on online users' information need. Some of the examples of such services include Netflix and Youtube for watching movies and videos, Google News and Yahoo! News for reading news articles, Amazon and Ebay for online shopping etc [6]. These services provide a huge amount of interesting information and generate a myriad of user behavior data as online users access the available resources.

Generally in any ecommerce sites or in search engine user will enter his required item of interest in search tab. Then the system will generate a number of relative results for each search. For the user to searching the entire results and selecting the required product will consume more time. Sometimes we may not get accurate results. Reason for not getting accurate results is that web sites are based entirely on matching user keywords against product title or main product description instead of full data set. To save user's time and to obtain better results the concept of recommender systems is evolved.

Recommender systems use the ratings and reviews of the user from user profiles and provide appropriate recommendations to the user. Recommending related products to the user helps in providing better shopping experience to him. Recommender systems have been become extremely common in recent years. The main objective of recommendation system is to provide filter relevant items or products from large datasets. It relies on similarity comparison. Product recommendations can be done in different ways. Product recommendations are classified into alternate products, complementary products and generic products. Recommendations for alternate products offer suggestions for products that are similar to the use's choice and could bought as an alternative or replacement of the actual product. For example if the user want to buy a Sony laptop then the system may recommend some other laptops which have same specification. Recommendations for complementary products offer suggestions for products that would complement, complete or go well with the actual selected product by the user. For example if a customer want to purchase a mobile then the system may recommend some complementary accessories like panels or earphones etc. Generic product recommendation does not depend on any product as a baseline. They can be generated without any knowledge of the user's past or current product preferences.

## **Recommendation System**

The system that predicts relevant products using robust filtering technique is known as Recommendation System (RS). The main objective of Recommendation System is to perform task to filter relevant products from the large databases. It much relies upon similarity prediction. Recommendation System (RS) can be categorized into three different types based upon application used; they are content based technique, collaboration technique, and hybrid technique.

### **Content Based Technique**

Content based filtering technique [1] [9] recommends items based on a comparison of items present in the user profile and the list of items available in the database. The description of each item is mentioned as some keywords and a set of descriptors. The user profile is created based on the interests of the customer on various areas. During the construction of user profile, the description of the interested products is also mentioned in terms of the set of descriptors. The recommendation process basically consists in matching up the attributes of the user profile against the attributes of a content object. The products are recommended based on the results.

In this filtering process, first the content analyzer extracts features such as keywords, concepts etc from unstructured text to produce a structured item representation and it will be stored in the repository of represented items. Secondly user's profile is constructed based on the user choices. The user's reactions are collected in some way and stored in feedback repository. The supervised learning algorithms are used to generate a predictive user profile, which is used for later filtering. Some potential strategy is implemented by the filtering component to rank the items present the database against user interested items mentioned in user profile. Among these the top ranked items are recommended to the user. Users interests will change from time to time so up-date data information must be maintained. To provide better recommendations the user profiles must be updated on regular basis. Feedback should be gathered on provided recommendation by asking the user whether the recommendations are satisfactory to the user or not.

Content based techniques provide complete user independence. Recommendations are made completely depending on the ratings provided by the user from user profiles. So ratings provided by an unknown person are not considered here. Explanations on how the recommender systems works can be provided by explicitly listening content feature or description that caused an item to occur in the list of recommendations this leads to the concept of transparency. They are capable of recommending items that are not yet rated by the user

Content based techniques are easy to implement but they have some shortcomings. In content based filtering there is a natural limit in number and type of feature associated with an item. No content based recommendation system provide suitable suggestions if the analyzed content does not contain enough information to determine items the user likes from the items user does not like. The system suggests items whose scores are high when matched against the user profile; hence the user is going to be recommended items, similar to those already rated. This drawback is also called 'serendipity' problem to highlight the tendency of content based systems to produce

recommendations with a limited degree of novelty. When there is insufficient description of the product it leads to 'sparsity' problem and the system is unable to recommend such products. When a new item or new user is added to the database it leads to cold start problem.

### **Collaboration Technique:**

Collaborative filtering [7] is one of the widely used techniques in recommendation systems. Application of collaborative filtering involves very large datasets because they can be applied to different kinds of data. The process of filtering for information or patterns using techniques involving collaboration among multiple agents, viewpoints, and data sources, etc. in collaborative filtering technique recommendations are made according to the preferences of other users who have similar interests as the customer [1]. It uses the choices of other users to make recommendations. The idea for collaborative filtering comes from the idea that people generally get the best recommendations from someone who has similar tastes as them. This model explores techniques for matching people with similar interests as user and making recommendations on this basis. With similar users, the system recommends items that the similar users have rated highly but not yet being rated by this user.

CF algorithms are required to have the ability to deal with highly sparse data, to scale with the more numbers of users and items, and to make satisfactory recommendations in a short time period. In early stages memory based collaborative filtering methods are used. This mechanism uses user rating data to compute similarity between users or items. They are easily to implement and highly effective and also there is no need to consider the content of items being recommended. It also have some limitations like, they are dependent on rating provided by the user. They are inefficient towards sparse data. The introduction of new users or new items can cause cold start problem, as there will be insufficient data on these new entries for the collaborative filtering. It also has limited scalability for large datasets.

To overcome the shortcomings of memory based collaborative filtering approach and to achieve better prediction performance 'model based' collaborative filtering approach is proposed. Here models are developed using data mining, machine learning algorithms to find patterns on training data. It uses the pure rating data to estimate or learn a model to make predictions. The sparsity, scalability and other problems are addressed effectively. It also improves prediction performance. The limitations of this model are, it is expensive to build and we have to lose useful information for dimensionality reduction techniques.

Hybrid CF techniques, such as the content-boosted CF algorithm and Personality Diagnosis (PD), combine CF and content-based techniques, avoid the limitations of either of above approaches and thereby improve performance of product recommendation. The advantage of this technique is it improves prediction performance. It effectively overcomes the problems such as sparsity and grey sheep. The limitations of this model are increased complexity and expense of implementation.

### **Hybrid Technique**

Though collaborative and content based techniques are widely used recommendation of products, they have some limitations and shortcomings. Both the models fail to provide effective recommendations in case of sparsity problem and cold start problem. So combining collaborative and content based filtering techniques can give better results. This leads to the concept of hybrid approach. Hybrid approaches can be implemented in several ways like combining the results of both collaborative and content based approaches and providing recommendations. Sometimes it combines the features of both the techniques and produces a single result set. These methods can overcome the problem like sparsity, cold start and grey sheep problem. The cold start problem is often reduced by a default rating to the new item based on the ratings assigned by the community to other similar items. . The term hybrid recommender system is used here to describe any recommender system that combines multiple recommendation techniques together to produce its output. There are seven hybridization methods:

- **Weighted:** The scores of several recommendation techniques are combined together to produce a single recommendation.
- **Switching:** The system switches between recommendations techniques depending on the current situation.
- **Mixed:** Recommendation system several recommendations are presented at the same time.
- **Feature combination:** The features of different recommendations data sources are thrown together in a single recommendation algorithms.
- **Feature argumentation:** output of one technique is used as input feature of another.
- **Cascade:** One recommender redefines the recommendation given by another.
- **Meta-level:** The model learned by one recommender is used an input to another.

### **Demographic Recommender Systems:**

Demographic recommender systems [1] [9] provide recommendations based on the demographic profile of the user. Recommended products can be produced for different demographic domains, by considering ratings of user in those respective domains. Constructing a demographic recommendation component is more difficult as it require access to the user's personal demographic data, which is not found in ratings and reviews of the user. Demographic information is used to identify the types of user that like a particular product. We use text classification to classify a user. Generally we will classify the data available in web pages and its related data. These methods attempt to find regularities among the description of the ser that like a particular product. It will categorize the user based on personal attributes and make recommendations based on demographic classes. The representation of the demographic information in a user model can vary greatly. It is similar to collaborative recommendation system. Demographic data forms people to people correlations like collaborative but uses different data. The main advantage of the recommender systems is it may not require the history of user rating because of which

it does not contain 'new user' problem [1]. Collaborative and demographic techniques that are discussed here have unique capacity to identify cross genre niches and can entice user to jump outside of the familiar.

### **Knowledge Based Recommendation Systems:**

Knowledge based recommendation systems [10] will make recommendations based on the interference about user needs and preferences. They have knowledge about how a particular item will meet particular user need and establishes relation between need and particular recommendation. For example in Google search engine when we type a certain keyword or sentence then it will recommend items based on the context and user need. Once user profile has been established in the system it is difficult to change the one's preferences. Due to this we may not able to provide accurate recommendations to the user. But, knowledge based recommenders will respond to the user's immediate need and they do not need any type of retraining when preferences change. The knowledge engineer should estimate the future recommendations of the user. So, knowledge engineering is required for this purpose. It can make recommendations as wide ranging as its knowledge base allows. It is sensitive to change of preferences and can include non product features. These types of recommender systems does not have 'ramp up' problem because they don't rely on historical data having user's preferences [1] [2]. But the drawback is knowledge acquisition bottlenecks triggered by the need of defining recommendation knowledge in an explicit fashion. Modeling, building and maintaining of the knowledge database is the main problem in knowledge based recommendation systems because it need a lot of effort and heavy human intervention. We have to maintain correct and up to date description of the items to be recommended. For this purpose qualitative as well as reliable data should be maintained. The advantage is there is no need of computing large amount of data to make recommendations. Most important thing is it does not contain 'cold start' problem. Whenever a new user or item is added with its description, the system is immediately able to compute recommendation for new user or item. They are able to generate proofs for those items which are recommended. Web of data is considered as an interesting source of information used by knowledge based recommender systems. Use of semantic web tools enhances the performance of knowledge database.

### **Utility Based Recommender Systems:**

Utility based recommendation systems make recommendations based on the computation of the utility of each item for a user. They use features of items as background data and derive utility functions over items from user to describe user preferences and apply the function to determine rank of function to user. The user must build complete preference function and weigh each attribute interaction. So it needs more user interaction. The central problem is how a utility function for each user should be created. It does not have ramp up and sparsity problem

**Table 1:** Different Recommendation Techniques

<b>Technique</b>	<b>Process</b>	<b>Advantages</b>	<b>Disadvantages</b>
<b>Collaborative</b>	Uses user-item ratings to make prediction	<ul style="list-style-type: none"> <li>• Quality improves over time</li> <li>• Domain knowledge is not needed</li> </ul>	<ul style="list-style-type: none"> <li>• Gray sheep problem</li> <li>• Cold start problem</li> <li>• Quality depends on large historical data set</li> </ul>
<b>Content based</b>	Relay on features of user and items and prediction	<ul style="list-style-type: none"> <li>• User independence</li> <li>• Implicit feedback is sufficient</li> </ul>	<ul style="list-style-type: none"> <li>• New user problem</li> <li>• Sparsity problem</li> </ul>
<b>Hybrid</b>	Combination of collaborative and content based technique	<ul style="list-style-type: none"> <li>• Cold start and sparsity problems can be reduced</li> </ul>	<ul style="list-style-type: none"> <li>• High complexity</li> <li>• High implementation cost</li> </ul>
<b>Demographic</b>	Identify users that are demographically similar to customer and recommend items depending on their ratings	<ul style="list-style-type: none"> <li>• Capacity to identify cross genre niches</li> <li>• Domain knowledge is not needed</li> <li>• No new-user problem</li> </ul>	<ul style="list-style-type: none"> <li>• Must gather demographic information</li> <li>• Gary sheep problem</li> <li>• Cold start problem</li> </ul>
<b>Knowledge based</b>	Matches user need for an item and recommends product according to it	<ul style="list-style-type: none"> <li>• No ramp up required</li> <li>• No cold start problem</li> <li>• Sensitive to change of preferences</li> </ul>	<ul style="list-style-type: none"> <li>• Knowledge engineering required</li> <li>• Suggestion ability is static</li> <li>• Heavy human intervention is needed</li> </ul>
<b>Utility based</b>	Applies utility functions over items that describes users preferences and recommend products based on rank of items	<ul style="list-style-type: none"> <li>• No data sparsity problem</li> <li>• No ramp up required</li> <li>• Can include non product features</li> <li>• Sensitive to change of preferences</li> </ul>	<ul style="list-style-type: none"> <li>• User must input utility function</li> <li>• Need more user interaction</li> </ul>

### **Conclusion and Future Work**

Recommender systems provide recommendation of products or services to users. Based on customers’ preferences or information needs, personalization can be achieved in these systems by comparing user profiles with items in a large item repository [6]. For recommending products to the user we have to consider several factors like information from user profiles, demographic information of the users etc. The recommendation technologies enable customers to obtain interesting information from the sources, as well as service providers to acquire remarkable economic benefit. Hence, recommender systems play different roles in satisfying the requirements of both customers and service providers. Even though the above mentioned recommendation techniques have some limitations they can be rectified through other techniques. The main problems observed in above described techniques are cold start

problem which will be obtained due to insufficient data in new entries. In recommender systems the cold start problem is often reduced by adopting a hybrid approach between collaborative and content based filtering techniques. Sparsity problem, new user and new item [1] problems are also addressed in above discussion. In future by combining different techniques we can provide better recommendations and also high performance.

## References

- [1] Robin Burke, "Hybrid Recommender Systems: Survey and Experiments", *User Modeling and User-Adapted Interaction* 12: 331-370, 2002.
- [2] Thomas Tran and Robin Cohen, "Hybrid Recommender Systems for Electronic Commerce", AAAI Technical Report WS-00-04.
- [3] J. Bobadilla, F. Ortega, A. Hernando, A. Gutiérrez, "Recommender systems survey", Elsevier, *Knowledge-Based Systems* 46 (2013) 109–132.
- [4] Emmanouil Vozalis, K.G. Margaritis, *Analysis of Recommender Systems' Algorithms*. Parallel and Distributed Processing Laboratory.
- [5] Baoyao Zhou, Siu C. Hui, Jie Tang, "Generation of Personalized Ontology Based on Consumer Emotion and Behavior Analysis", *IEEE Transactions On Affective Computing*, Vol. 3, No. 2, April-June 2012.
- [6] Li, Lei, "Next Generation of Recommender Systems: Algorithms and Applications" (2014). FIU Electronic Theses and Dissertations, Paper 1446.
- [7] Jonathan L. Herlocker , Joseph A. Konstan, Loren G. Terveen, And John T. Riedl, " Evaluating Collaborative Filtering Recommender Systems ", *ACM Transactions on Information Systems (TOIS)* , Volume 22, Issue 1, January 2004, Page 5-53.
- [8] Felix Gao, Calvin Lei, "Movie Recommendation system".
- [9] Michael J. Pazzani, "A Framework for Collaborative, Content-Based and Demographic Filtering", *Artificial Intelligence Review*, December 1999, Volume 13, Issue 5-6, pp 393-408.
- [10] Robin Burke, "Knowledge-based recommender systems".