

## **Introduction:**

The explosive growth in the amount of available digital information and the number of visitors to the Internet have created a potential challenge of information overload which hinders timely access to items of interest on the Internet. Information retrieval systems, such as Google, DevilFinder and Altavista have partially solved this problem but prioritization and personalization (where a system maps available content to user's interests and preferences) of information were absent. This has increased the demand for recommender systems more than ever before. Recommender systems are information filtering systems that deal with the problem of information overload by filtering vital information fragment out of large amount of dynamically generated information according to user's preferences, interest, or observed behavior about item . Recommender system has the ability to predict whether a particular user would prefer an item or not based on the user's profile.

Recommender systems are beneficial to both service providers and users . They reduce transaction costs of finding and selecting items in an online shopping environment . Recommendation systems have also proved to improve decision making process and quality . In e-commerce setting, recommender systems enhance revenues, for the fact that they are effective means of selling more products. In scientific libraries, recommender systems support users by allowing them to move beyond catalog searches. Therefore, the need to use efficient and accurate recommendation techniques within a system that will provide relevant and dependable recommendations for users cannot be over-emphasized.

## **Literature Review:**

Recommender system is defined as a decision making strategy for users under complex information environments . Also, recommender system was defined from the perspective of E-commerce as a tool that helps users search through records of knowledge which is related to users' interest and preference . Recommender system was defined as a means of assisting and augmenting the social process of using recommendations of others to make choices when there is no sufficient personal knowledge or experience of the alternatives . Recommender systems handle the problem of information overload that users normally encounter by providing them with personalized, exclusive content and service recommendations. Recently, various approaches for building recommendation systems have been developed, which can utilize either collaborative filtering, content-based filtering or hybrid filtering . Collaborative filtering technique is the most mature and the most commonly implemented. Collaborative filtering recommends items by identifying other users with similar taste; it uses their opinion to recommend items to the active user. Collaborative recommender systems have been implemented in different application areas. GroupLens is a news-based architecture which employed collaborative methods in assisting users to locate articles from massive news database . Ringo is an online social information filtering system that uses collaborative filtering to build users profile based on their ratings on music albums . Amazon uses topic diversification algorithms to improve its recommendation . The system uses collaborative filtering method to overcome scalability issue by generating a table of similar items offline through the use of item-to-item matrix. The system then recommends other products which are similar online according to the users' purchase history. On the other hand, content-based techniques match content resources to user characteristics. Content-based filtering techniques normally base their predictions on user's information, and they ignore contributions from other users as with the case of collaborative techniques . Fab relies heavily on the ratings of different users in order to create a training set and it is an example of content-based recommender system. Some other systems that use content-based filtering to help users find information on the Internet include Letizia . The system makes use of a user interface that assists users in browsing the Internet; it is able to track the browsing pattern of a user to predict the pages that they may be interested in. Pazzani et al. designed an intelligent agent that attempts to predict which web pages will interest a user by using naive Bayesian classifier. The agent allows a user to provide training instances by rating different pages as either hot or cold. Jennings and Higuchi describe a neural network that models the interests of a user in a Usenet news environment.

Despite the success of these two filtering techniques, several limitations have been identified. Some of the problems associated with content-based filtering techniques are limited content analysis, overspecialization and sparsity of data . Also, collaborative approaches exhibit cold-start, sparsity and scalability problems. These problems usually reduce the quality of recommendations. In order to mitigate some of the problems identified, Hybrid filtering, which combines two or more filtering techniques in different ways in order to increase the accuracy and performance of recommender systems has been proposed . These techniques combine two or more filtering approaches in order to harness their strengths while leveling out their corresponding weaknesses . They can be classified based on their operations into weighted hybrid, mixed hybrid, switching hybrid, feature-combination hybrid, cascade hybrid, feature-augmented hybrid and meta-level hybrid . Collaborative filtering and content-based filtering approaches are widely used today by implementing content-based and collaborative techniques differently and the results of their prediction later combined or adding the characteristics of content-based to collaborative filtering and vice versa. Finally, a general unified model which incorporates both content-based and collaborative filtering properties could be developed . The problem of sparsity of data and cold-start was addressed by combining the ratings, features and demographic information about items in a cascade hybrid recommendation technique in . In Ziegler et al. , a hybrid collaborative filtering approach was proposed to exploit bulk taxonomic information designed for exacting product classification to address the data sparsity problem of CF recommendations, based on the generation of profiles via inference of super-topic score and topic diversification. A hybrid recommendation technique is also proposed in Ghazantar and Pragel-Benett , and this uses the content-based profile of individual user to find similar users which are used to make predictions. In Sarwar et al. , collaborative filtering was combined with an information filtering agent. Here, the authors proposed a framework for integrating the content-based filtering agents and collaborative filtering. A hybrid recommender algorithm is employed by many applications as a result of new user problem of content-based filtering techniques and average user problem of collaborative filtering . A simple and straightforward method for combining content-based and collaborative filtering was proposed by Cunningham et al. . A music recommendation system which combined tagging information, play counts and social relations was proposed in Konstas et al. . In order to determine the number of neighbors that can be automatically connected on a social platform, Lee and Brusilovsky embedded social information into collaborative filtering algorithm. A Bayesian mixed-effects model that integrates user ratings, user and item features in a single unified framework was proposed by Condifff et al..

## **Objective:**

On the Internet, where the number of choices is overwhelming, there is need to filter, prioritize and efficiently deliver relevant information in order to alleviate the problem of information overload, which has created a potential problem to many Internet users. Recommender systems solve this problem by searching through large volume of dynamically generated information to provide users with personalized content and services. This paper explores the different characteristics and potentials of different prediction techniques in recommendation systems in order to serve as a compass for research and practice in the field of recommendation systems.

## **System Design:**

### **Phases of recommendation process**

#### **1. Information collection phase:**

This collects relevant information of users to generate a user profile or model for the prediction tasks including user's attribute, behaviors or content of the resources the user accesses. A recommendation agent cannot function accurately until the user profile/model has been well constructed. The system needs to know as much as possible from the user in order to provide reasonable recommendation right from the onset. Recommender systems rely on different types of input such as the most convenient high quality explicit feedback, which includes explicit input by users regarding their interest in item or implicit feedback by inferring user preferences indirectly through observing user behavior. Hybrid feedback can also be obtained through the combination of both explicit and implicit feedback. In E-learning platform, a user profile is a collection of personal information associated with a specific user. This information includes cognitive skills, intellectual abilities, learning styles, interest, preferences and interaction with the system. The user profile is normally used to retrieve the needed information to build up a model of the user. Thus, a user profile describes a simple user model. The success of any recommendation system depends largely on its ability to represent user's current interests. Accurate models are indispensable for obtaining relevant and accurate recommendations from any prediction techniques.

#### **2. Explicit feedback:**

The system normally prompts the user through the system interface to provide ratings for items in order to construct and improve his model. The accuracy of recommendation depends on the quantity of ratings provided by the user. The only shortcoming of this method is, it requires effort from the users and also, users are not always ready to supply enough information. Despite the fact that explicit feedback requires more effort from user, it is still seen as providing more reliable data, since it does not involve extracting preferences from actions, and it also provides transparency into the recommendation process that results in a slightly higher perceived recommendation quality and more confidence in the recommendations .

### **3. Implicit feedback:**

The system automatically infers the user's preferences by monitoring the different actions of users such as the history of purchases, navigation history, and time spent on some web pages, links followed by the user, content of e-mail and button clicks among others. Implicit feedback reduces the burden on users by inferring their user's preferences from their behavior with the system. The method though does not require effort from the user, but it is less accurate. Also, it has also been argued that implicit preference data might in actuality be more objective, as there is no bias arising from users responding in a socially desirable way and there are no self-image issues or any need for maintaining an image for others .

### **4. Hybrid feedback:**

The strengths of both implicit and explicit feedback can be combined in a hybrid system in order to minimize their weaknesses and get a best performing system. This can be achieved by using an implicit data as a check on explicit rating or allowing user to give explicit feedback only when he chooses to express explicit interest.

### **5. Learning phase:**

It applies a learning algorithm to filter and exploit the user's features from the feedback gathered in information collection phase.

### **6. Prediction/recommendation phase:**

It recommends or predicts what kind of items the user may prefer. This can be made either directly based on the dataset collected in information collection phase which could be memory based or model based or through the system's observed activities of the user.

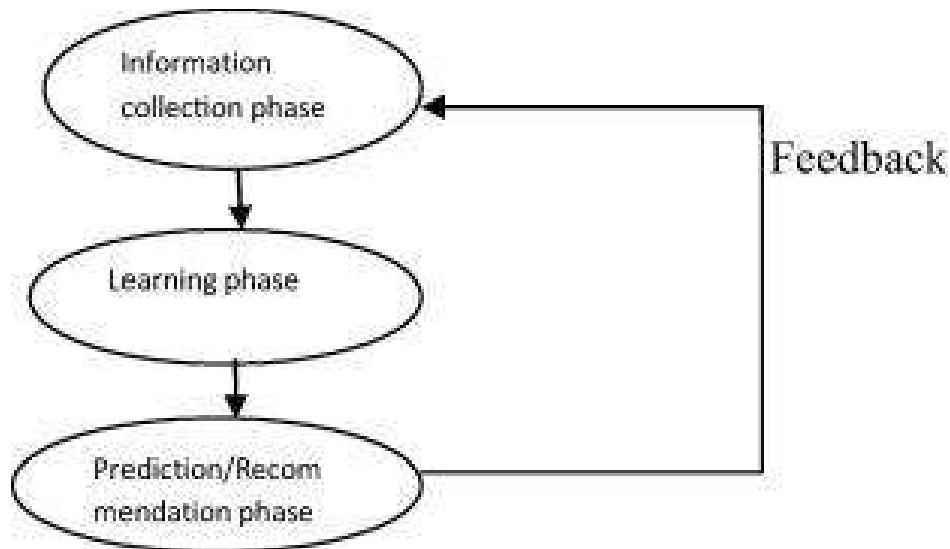


Fig:1

### **Recommendation filtering techniques:**

The use of efficient and accurate recommendation techniques is very important for a system that will provide good and useful recommendation to its individual users. This explains the importance of understanding the features and potentials of different recommendation techniques. Below figure shows the anatomy of different recommendation filtering techniques.

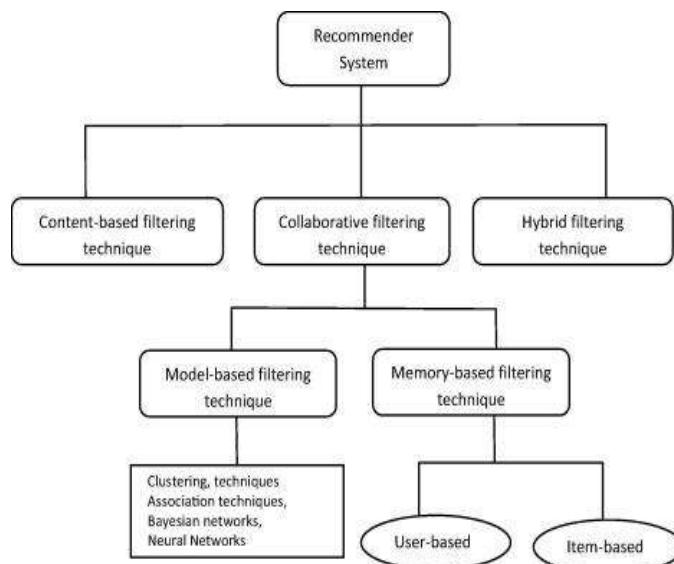


Fig:2

### **1. Content-based filtering:**

Content-based technique is a domain-dependent algorithm and it emphasizes more on the analysis of the attributes of items in order to generate predictions. When documents such as web pages, publications and news are to be recommended, content-based filtering technique is the most successful. In content-based filtering technique, recommendation is made based on the user profiles using features extracted from the content of the items the user has evaluated in the past . Items that are mostly related to the positively rated items are recommended to the user. CBF uses different types of models to find similarity between documents in order to generate meaningful recommendations. It could use Vector Space Model such as Term Frequency Inverse Document Frequency (TF/IDF) or Probabilistic models such as Naïve Bayes Classifier , Decision Trees or Neural Networks to model the relationship between different documents within a corpus. These techniques make recommendations by learning the underlying model with either statistical analysis or machine learning techniques. Content-based filtering technique does not need the profile of other users since they do not influence recommendation. Also, if the user profile changes, CBF technique still has the potential to adjust its recommendations within a very short period of time. The major disadvantage of this technique is the need to have an in-depth knowledge and description of the features of the items in the profile.

### **Pros and Cons of content-based filtering techniques:**

CB filtering techniques overcome the challenges of CF. They have the ability to recommend new items even if there are no ratings provided by users. So even if the database does not contain user preferences, recommendation accuracy is not affected. Also, if the user preferences change, it has the capacity to adjust its recommendations in a short span of time. They can manage situations where different users do not share the same items, but only identical items according to their intrinsic features. Users can get recommendations without sharing their profile, and this ensures privacy . CBF technique can also provide explanations on how recommendations are generated to users. However, the techniques suffer from various problems as discussed in the literature . Content based filtering techniques are dependent on items' metadata. That is, they require rich description of items and very well organized user profile before recommendation can be made to users. This is called limited content analysis. So, the effectiveness of CBF depends on the availability of descriptive data. Content overspecialization is another serious problem of CBF technique. Users are restricted to getting recommendations similar to items already defined in their profiles.

### **Examples of content-based filtering systems:**

News Dude is a personal news system that utilizes synthesized speech to read news stories to users. TF-IDF model is used to describe news stories in order to determine the short-term recommendations which is then compared with the Cosine Similarity Measure and finally supplied to a learning algorithm (NN). CiteSeer is an automatic citation indexing that uses various heuristics and machine learning algorithms to process documents. Today, CiteSeer is among the largest and widely used research paper repository on the web.

LIBRA is a content-based book recommendation system that uses information about book gathered from the Web. It implements a Naïve Bayes classifier on the information extracted from the web to learn a user profile to produce a ranked list of titles based on training examples supplied by an individual user. The system is able to provide explanation on any recommendations made to users by listing the features that contribute to the highest ratings and hence allowing the users to have total confidence on the recommendations provided to users by the system.

## **2. Collaborative filtering:**

Collaborative filtering is a domain-independent prediction technique for content that cannot easily and adequately be described by metadata such as movies and music. Collaborative filtering technique works by building a database (user-item matrix) of preferences for items by users. It then matches users with relevant interest and preferences by calculating similarities between their profiles to make recommendations . Such users build a group called neighborhood. An user gets recommendations to those items that he has not rated before but that were already positively rated by users in his neighborhood. Recommendations that are produced by CF can be of either prediction or recommendation. Prediction is a numerical value,  $R_{ij}$ , expressing the predicted score of item j for the user i, while Recommendation is a list of top N items that the user will like the most as shown in below figure. The technique of collaborative filtering can be divided into two categories: memory-based and model-based .

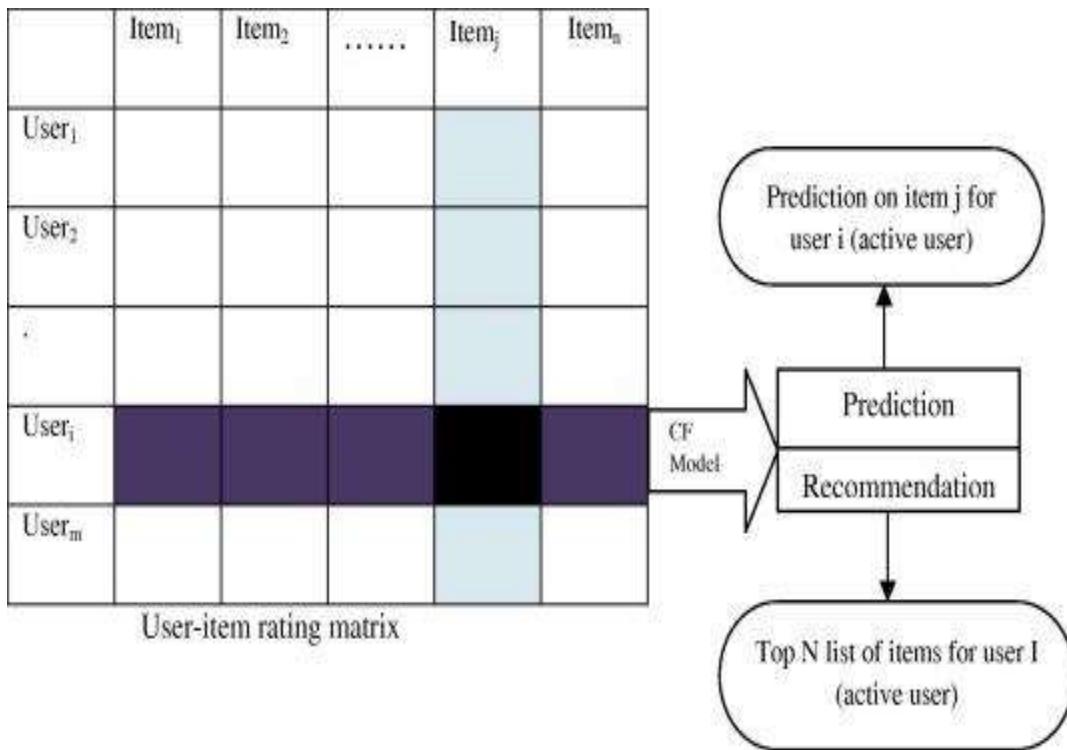


Fig:3

## 2.1. Memory based techniques:

The items that were already rated by the user before play a relevant role in searching for a neighbor that shares appreciation with him. Once a neighbor of a user is found, different algorithms can be used to combine the preferences of neighbors to generate recommendations. Due to the effectiveness of these techniques, they have achieved widespread success in real life applications. Memory-based CF can be achieved in two ways through user-based and item-based techniques. User based collaborative filtering technique calculates similarity between users by comparing their ratings on the same item, and it then computes the predicted rating for an item by the active user as a weighted average of the ratings of the item by users similar to the active user where weights are the similarities of these users with the target item. Item-based filtering techniques compute predictions using the similarity between items and not the similarity between users. It builds a model of item similarities by retrieving all items rated by an active user from the user-item matrix, it determines how similar the retrieved items are to the target item, then it selects the k most similar items and their corresponding similarities are also determined.

Prediction is made by taking a weighted average of the active users rating on the similar items k. Several types of similarity measures are used to compute similarity between item/user. The two most popular similarity measures are correlation-based and cosine-based. Pearson correlation coefficient is used to measure the extent to which two variables linearly relate with each other and is defined as

$$s(a,u) = \frac{\sum_{i=1}^n (r_{a,i} - \bar{r}_a)(r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i=1}^n (r_{a,i} - \bar{r}_a)^2} \sqrt{\sum_{i=1}^n (r_{u,i} - \bar{r}_u)^2}}$$

From the above equation,

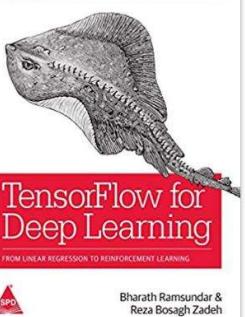
$S(a,u)$  is the mean rating given by user  $a$  while  $n$  is the total number of items in the user-item space. Also, prediction for an item is made from the weighted combination of the selected neighbors' ratings, which is computed as the weighted deviation from the neighbors' mean. The general prediction formula is

$$p(a,i) = \bar{r}_a + \frac{\sum_{i=1}^n (r_{u,i} - \bar{r}_u) \times s(a,u)}{\sum_{i=1}^n s(a,u)}$$

Cosine similarity is different from Pearson-based measure in that it is a vector-space model which is based on linear algebra rather than statistical approach. It measures the similarity between two  $n$ -dimensional vectors based on the angle between them. Cosine-based measure is widely used in the fields of information retrieval and texts mining to compare two text documents, in this case, documents are represented as vectors of terms. The similarity between two items  $u$  and  $v$  can be defined as

$$s(\vec{u}, \vec{v}) = \frac{\vec{u} \cdot \vec{v}}{|\vec{u}| * |\vec{v}|} = \frac{\sum_i r_{u,i} r_{v,i}}{\sqrt{\sum_i r_{u,i}^2} \times \sqrt{\sum_i r_{v,i}^2}}$$

Similarity measure is also referred to as similarity metric, and they are methods used to calculate the scores that express how similar users or items are to each other. These scores can then be used as the foundation of user- or item-based recommendation generation. Depending on the context of use, similarity metrics can also be referred to as correlation metrics or distance metrics.



TensorFlow for Deep Learning  
FROM LINEAR REGRESSION TO REINFORCEMENT LEARNING

by BharathRamsundar (Author), Reza Zadeh (Author)  
Be the first to review this item

See all formats and editions

Paperback  
750.00

2 New from 750.00

Delivery to pincode 700010 - Kolkata within 2 - 4 business days. Details

Report incorrect product information.

See this image

**Special offers and product promotions**

- Shop with balance and get 10% cashback, up to Rs.300 per customer. Minimum order Rs.250 (excluding delivery fees). Applicable on orders paid using only Amazon Pay balance during May 13-16. ADD BALANCE Here's how (terms and conditions apply)
- 10% Cashback with ICICI Bank cards. Here's how (terms and conditions apply)

**Customers who viewed this item also viewed**

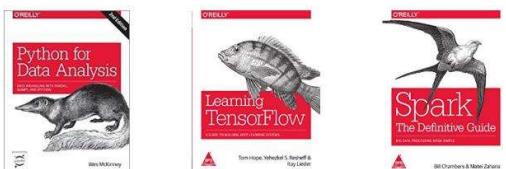


Fig:4

## **2.2. Model-based techniques:**

This technique employs the previous ratings to learn a model in order to improve the performance of Collaborative filtering Technique. The model building process can be done using machine learning or data mining techniques. These techniques can quickly recommend a set of items for the fact that they use pre-computed model and they have proved to produce recommendation results that are similar to neighborhood-based recommender techniques.

Examples of these techniques include Dimensionality Reduction technique such as Singular Value Decomposition (SVD), Matrix Completion Technique, Latent Semantic methods, and Regression and Clustering. Model-based techniques analyze the user-item matrix to identify relations between items; they use these relations to compare the list of top-N recommendations. Model based techniques resolve the sparsity problems associated with recommendation systems.

The use of learning algorithms has also changed the manner of recommendations from recommending what to consume by users to recommending when to actually consume a product. It is therefore very important to examine other learning algorithms used in model-based recommender systems:

**Association rule:** Association rules mining algorithms extract rules that predict the occurrence of an item based on the presence of other items in a transaction. For instance, given a set of transactions, where each transaction is a set of items, an association rule applies the form  $A \rightarrow B$ , where A and B are two sets of items . Association rules can form a very compact representation of preference data that may improve efficiency of storage as well as performance. Also, the effectiveness of association rule for uncovering patterns and driving personalized marketing decisions has been known for sometimes . However, there is a clear relation between this method and the goal of a Recommendation System but they have not become mainstream.

**Clustering:** Clustering techniques have been applied in different domains such as, pattern recognition, image processing, statistical data analysis and knowledge discovery. Clustering algorithm tries to partition a set of data into a set of sub-clusters in order to discover meaningful groups that exist within them. Once clusters have been formed, the opinions of other users in a cluster can be averaged and used to make recommendations for individual users. A good clustering method will produce high quality clusters in which the intra-cluster similarity is high, while the inter-cluster similarity is low. In some clustering approaches, a user can have partial participation in different clusters, and recommendations are then based on the average across the clusters of participation which is weighted by degree of participation. K-means and Self-Organizing Map (SOM) are the most commonly used among the different clustering methods. K-means takes an input parameter, and then partitions a set of n items into K clusters . The Self-Organizing Map (SOM) is a method for an unsupervised learning, based on artificial neurons clustering technique. Clustering techniques can be used to reduce the candidate set in collaborative-based algorithms.

**Decision tree:** Decision tree is based on the methodology of tree graphs which is constructed by analyzing a set of training examples for which the class labels are known. They are then applied to classify previously unseen examples. If trained on very high quality data, they have the ability to make very accurate predictions. Decision trees are more interpretable than other classifier such as Support Vector machine (SVM) and Neural Networks because they combine simple questions about data in an understandable manner. Decision trees are also flexible in handling items with mixture of real-valued and categorical features as well as items that have some specific missing features.

**Artificial Neural network:** ANN is a structure of many connected neurons (nodes) which are arranged in layers in systematic ways. The connections between neurons have weights associated with them depending on the amount of influence one neuron has on another. There are some advantages in using neural networks in some special problem situations. For example, due to the fact that it contains many neurons and also assigned weight to each connection, an artificial neural network is quite robust with respect to noisy and erroneous data sets. ANN has the ability of estimating nonlinear functions and capturing complex relationships in data sets also, they can be efficient and even operate if part of the network fails. The major disadvantage is that it is hard to come up with the ideal network topology for a given problem and once the topology is decided this will act as a lower bound for the classification error.

**Link analysis:** Link Analysis is the process of building up networks of interconnected objects in order to explore pattern and trends. It has presented great potentials in improving the accomplishment of web search. Link analysis consists of PageRank and HITS algorithms. Most link analysis algorithms handle a web page as a single node in the web graph.

**Regression:** Regression analysis is used when two or more variables are thought to be systematically connected by a linear relationship. It is a powerful and diversity process for analyzing associative relationships between dependent variable and one or more independent variables. Uses of regression contain curve fitting, prediction, and testing systematic hypotheses about relationships between variables. The curve can be useful to identify a trend within dataset, whether it is linear, parabolic, or of some other forms.

**Bayesian Classifiers:** They are probabilistic framework for solving classification problems which is based on the definition of conditional probability and Bayes theorem. Bayesian classifiers consider each attribute and class label as random variables. Given a record of N features ( $A_1, A_2, \dots, A_N$ ), the goal of the classifier is to predict class  $C_k$  by finding the value of  $C_k$  that maximizes the posterior probability of the class given the data  $P(C_k|A_1, A_2, \dots, A_N)$  by applying Bayes' theorem,  $P(C_k|A_1, A_2, \dots, A_N) \propto P(A_1, A_2, \dots, A_N|C_k)P(C_k)$ . The most commonly used Bayesian classifier is known as the Naive Bayes Classifier. In order to estimate the conditional probability,  $P(A_1, A_2, \dots, A_N|C_k)$ , a Naive Bayes Classifier assumes the probabilistic independence of the attributes that is, the presence or absence of a particular attribute is unrelated to the presence or absence of any other. This assumption leads to  $P(A_1, A_2, \dots, A_N|C_k) = P(A_1|C_k)P(A_2|C_k)\dots P(A_N|C_k)$ .

....,  $A_N|C_k) = P(A_1|C_k)P(A_2|C_k)\dots P(A_N|C_k)$ . The main benefits of Naive Bayes classifiers are that they are robust to isolated noise points and irrelevant attributes, and they handle missing values by ignoring the instance during probability estimate calculations. However, the independence assumption may not hold for some attributes as they might be correlated. In this case, the usual approach is to use Bayesian Networks. Bayesian classifiers may prove practical for environments in which knowledge of user preferences changes slowly with respect to the time needed to build the model but are not suitable for environments in which users preference models must be updated rapidly or frequently. It is also successful in model-based recommendation systems because it is often used to derive a model for content-based recommendation systems.

**Matrix completion techniques:** The essence of matrix completion technique is to predict the unknown values within the user-item matrices. Correlation based K-nearest neighbor is one of the major techniques employed in collaborative filtering recommendation systems. They depend largely on the historical rating data of users on items. Most of the time, the rating matrix is always very big and sparse due to the fact that users do not rate most of the items represented within the matrix. This problem always leads to the inability of the system to give reliable and accurate recommendations to users. Different variations of low rank models have been used in practice for matrix completion especially toward application in collaborative filtering . Formally, the task of matrix completion technique is to estimate the entries of a matrix,  $M \in R^{m \times n}$ , when a subset,  $\Omega \subset \{(i,j) : 1 \leq i \leq m, 1 \leq j \leq n\}$

of the new entries is observed, a particular set of low rank matrices,

$M^{\hat{}} = U V^T$ , where  $U \in R^{m \times k}$  and  $V \in R^{k \times n}$  and  $k \ll \min(m,n)$

. The most widely used algorithm in practice for recovering  $M$  from partially observed matrix using low rank assumption is Alternating Least Square (ALS) minimization which involves optimizing over  $U$  and  $V$  in an alternating manner to minimize the square error over observed entries while keeping other factors fixed. Candes and Recht proposed the use of matrix completion technique in the Netflix problem as a practical example for the utilization of the technique. Keshavan et al. used SVD technique in an OptSpace algorithm to deal with matrix completion problem. The result of their experiment showed that SVD is able provide a reliable initial estimate for spanning subspace which can be further refined by gradient descent on a Grassmannian manifold. Model based techniques solve sparsity problem. The major drawback of the techniques is that the model building process is computationally expensive and the capacity of memory usage is highly intensive. Also, they do not alleviate the cold-start problem.

### **Pros and Cons of collaborative filtering techniques:**

Collaborative Filtering has some major advantages over CBF in that it can perform in domains

where there is not much content associated with items and where content is difficult for a computer system to analyze (such as opinions and ideal). Also, CF technique has the ability to provide serendipitous recommendations, which means that it can recommend items that are relevant to the user even without the content being in the user's profile. Despite the success of CF techniques, their widespread use has revealed some potential problems such as follows.

### **Cold-start problem:**

This refers to a situation where a recommender does not have adequate information about a user or an item in order to make relevant predictions. This is one of the major problems that reduce the performance of recommendation system. The profile of such new user or item will be empty since he has not rated any item; hence, his taste is not known to the system.

### **Data sparsity problem:**

This is the problem that occurs as a result of lack of enough information, that is, when only a few of the total number of items available in a database are rated by users. This always leads to a sparse user-item matrix, inability to locate successful neighbors and finally, the generation of weak recommendations. Also, data sparsity always leads to coverage problems, which is the percentage of items in the system that recommendations can be made .

### **Scalability:**

This is another problem associated with recommendation algorithms because computation normally grows linearly with the number of users and items . A recommendation technique that is efficient when the number of dataset is limited may be unable to generate satisfactory number of recommendations when the volume of dataset is increased. Thus, it is crucial to apply recommendation techniques which are capable of scaling up in a successful manner as the number of dataset in a database increases. Methods used for solving scalability problem and speeding up recommendation generation are based on Dimensionality reduction techniques, such as Singular Value Decomposition (SVD) method, which has the ability to produce reliable and efficient recommendations.

### **Synonymy:**

Synonymy is the tendency of very similar items to have different names or entries. Most recommender systems find it difficult to make distinction between closely related items such as the difference between e.g. baby wear and baby cloth. Collaborative Filtering systems usually find no match between the two terms to be able to compute their similarity. Different methods, such as automatic term expansion, the construction of a thesaurus, and Singular Value Decomposition (SVD), especially Latent Semantic Indexing are capable of solving the synonymy problem. The shortcoming of these methods is that some added terms may have different meanings from what is intended, which sometimes leads to rapid degradation of

recommendation performance.

### **Examples of collaborative systems:**

**Ringo** is a user-based CF system which makes recommendations of music albums and artists. In Ringo, when a user initially enters the system, a list of 125 artists is given to the user to rate according to how much he likes listening to them. The list is made up of two different sections. The first session consists of the most often rated artists, and this affords the active user opportunity to rate artists which others have equally rated, so that there is a level of similarities between different users' profiles. The second session is generated upon a random selection of items from the entire user-item matrix, so that all artists and albums are eventually rated at some point in the initial rating phases.

**GroupLens** is a CF system that is based on client/server architecture; the system recommends Usenet news which is a high volume discussion list service on the Internet. The short lifetime of Netnews, and the underlying sparsity of the rating matrices are the two main challenges addressed by this system. Users and Netnews are clustered based on the existing news groups in the system, and the implicit ratings are computed by measuring the time the users spend reading Netnews.

**Amazon.com** is an example of e-commerce recommendation engine that uses scalable item-to-item collaborative filtering techniques to recommend online products for different users. The computational algorithm scales independently of the number of users and items within the database. Amazon.com uses an explicit information collection technique to obtain information from users. The interface is made up of the following sections, your browsing history, rate these items, and improve your recommendations and your profile. The system predicts users interest based on the items he/she has rated. The system then compares the users browsing pattern on the system and decides the item of interest to recommend to the user. Amazon.com popularized feature of “people who bought this item also bought these items”. Example of Amazon.com item-to-item contextual recommendation interface is shown in below figure