CHAPTER 2
LITERATURE SURVEY

1. Manos Papagelis, Dimitris Plexousakis (2005), “Qualitative Analysis of User-based and Item-based Prediction Algorithms for Recommendation Agents” In this paper, prediction algorithms are utilized to user-based and item-based similarity measures derived from either explicit or implicit ratings. The utilization of explicit ratings is an “implicit” sense which is to enrich a user’s model without actually prompting users to express their preference to Categories in similarity prediction. The formation of a range of item-based and user-based prediction algorithms are according to item-based and user-based similarity measures. Both statistical and decision-support accuracy metrics of the algorithms against different levels of data sparsity and different operational thresholds provide users with items that match their interests. Prediction algorithms combine user based and item based similarity measures derived from explicit or implicit ratings are to provide the recommendation for an item that matches with the user interests. Category-boosted predictions can lead to slightly better predictions when combined with explicit ratings, while implicit ratings perform much worse than explicit ratings.

2. Sung Young Jung; Jeong-Hee Hong; Taek-Soo Kim (2005), “A Statistical Model for User Preference”, This paper focuses on modelling user preference issues which is resolved through intelligent information systems. Automatically analyzing models can be classified into one of the following three classes, depending on how they represent preference: vector similarity, probability, and association. First, vector similarity is adopted in both collaborative filtering and
content-based filtering. In collaborative filtering, user preference to an item is represented by how many users with preferences similar to the given user choose the item. Second, probability is used to predict user’s future behaviours in a Bayesian network where preference is represented by the probability that a user selects for a given item. Third, in association rule mining Area preference is represented by the strength of association (i.e., correlation) between an item and a user history. In this literature, extensive research has been performed to automatically analyze user preference and to utilize it. The representation of preference, usually given by measure of vector similarity or probability, does not always correspond to common sense of preference to extract the rating of the user to the particular item. The preference can be classified into positive and negative preference which can be considered as feedback based explicit collaborative filtering model. The preference model using mutual information in combined information of joint features alleviates problems arising from sparse data.

3. Jun Wang, Arjen P. de Vries, Marcel J.T. Reinders(2006), “Unifying User based and Item based Collaborative Filtering Approaches by Similarity Fusions”. Here, proposed Memory based algorithms has been employed for predicting the new average ratings of similar users or items. Collaborative filtering predicts new ratings by averaging (weighted) ratings between, respectively, pairs of similar users or items. Large number of data were not available so the prediction quality may be poor. memory-based collaborative filtering problem in a generative probabilistic framework, treating individual user-item ratings as predictors of missing ratings. The final rating is estimated
by fusing predictions from three sources: predictions based on ratings of the
same item by other users, predictions based on different item ratings made by
the same user, and, third, ratings predicted based on data from other but similar
users rating other but similar items. Existing user-based and item-based
approaches correspond to the two simple cases of our framework. The complete
model is however more robust to data sparsity, because the different types of
ratings are used in concert, while additional ratings from similar users towards
similar items are employed as a background model to smooth the predictions.
Additional ratings from similar users for similar items have been considered to
smooth the predictions.

4. Nachiketa Sahoo, Ramayya Krishnan, George Duncan, James P.
Callan(2007), “Collaborative Filtering with Multi-component Rating for
Recommender Systems”. This work deals with multi-component rating was
considered as a better recommendation technique than single component
recommendation technique over the small training data. The dependency
structure among the component ratings is discovered and incorporated into a
mixture model. The parameters of the model were estimated using the
Expectation Maximization algorithm. It was found that using multiple
components leads to improved recommendations over using only one
component when small amount of training data is used. However, when enough
training data has no gain from using additional components of the ratings was
observed.

Facets”. In this paper, the importance of dominance and skyline analysis has
been well recognized in multi-criteria decision making applications. Many studies assume a fixed order on the attributes. In practice, different customers may have different preferences on nominal attributes. In this literature, an interesting data mining problem is identified, finding favourable facets, which has not been studied before. Given a set of points in a multidimensional space, for a specific target point \( p \) wants to discover with respect to which combinations of orders (e.g., customer preferences) on the nominal attributes \( p \) is not dominated by any other points. Such combinations are called the favorable facets of \( p \). Consider both the effectiveness and the efficiency of the mining in the work. A given point may have many favorable facets. The notion of minimal disqualifying condition (MDC) which is effective in summarizing favorable facets is proposed with development of efficient algorithms for favourable facet mining for different application scenarios. The first method computes favorable facets on the fly. The second method pre-computes all minimal disqualifying conditions so that the favourable facets can be looked up in constant time. An extensive performance study using both synthetic and real data sets is reported to verify their effectiveness and efficiency.

6. **Liang He, Weiwei Xia, Lei Ren (2008), “A Collaborative Filtering Algorithm based on Users Partial Similarity”**. This paper focuses on two main problems of Collaborative Filtering (CF): Scalability and Sparsity. CF problems have been handled by proposing collaborative filtering algorithm based on Users' Partial Similarity (CFUPS). Collaborative filtering problems on two crucial steps: (1) computing neighbor users for active users and (2) missing data prediction algorithm. CFUPS’s main idea is that to compute the similarity between users
who rely on partial items with their common interests, not on all common rated items and by combining items' attributes similarity and their ratings similarity properly for computing missing ratings. Theoretically, this method is effective in improving the recommendation precision and withstanding data sparsity.

7. **Song Jie Gong(2009), “A Collaborative Recommender Based on User Information and Item Information”**. In traditional CF, the similarity value between the target user and end user has been computed based on ratings information whereas hybrid CF which considers both user attribute information and item attribute information. Collaborative methods recommend items based on aggregated user ratings of those items and these techniques do not depend on the availability of textual descriptions. They share the common goal of assisting the user’s search for items of interest, and thus attempt to address one of the key research problems of the information age: locating needles in a haystack that is growing exponentially. Collaborative filtering systems can deal with large numbers of people and with many different items. User attribute information associated with a user's personality and item attribute information associated with an item's inside are rarely considered in the collaborative filtering recommendation process. Collaborative filtering approaches compute a similarity value between the target user and each other user by computing the relativity of their ratings, and they only consider the ratings information. A hybrid collaborative filtering recommender which employs the user attribute information and the item attribute information is modelled. This approach combines the user rating similarity and the user attribute similarity in the user based collaborative filtering process and then it combines the item rating
similarity and the item attribute similarity in the item based collaborative filtering process to produce recommendations. The collaborative filtering recommender employs the user attribute and item attribute can alleviate the sparsity issue in the recommender systems. This literature is suitable for item similarity formulation and user similarity formulation.

8. Jin Soung Yoo, Shashi Shekhar(2009), “Similarity-Profiled Temporal Association Mining”. This paper deals with Similarity-profiled temporal association mining is to discover all associated item sets whose prevalence variations over time are similar to the reference sequence under a threshold. Similarity-profiled temporal association mining can reveal interesting relationships of data items that co-occur with a particular event over time. Most works in temporal association mining have focused on capturing special temporal regulation patterns such as cyclic patterns and calendar scheme-based patterns of the user or item information. Association rules discover interrelationships among various data items in transactional data. However, the proposed model is flexible in representing interesting temporal patterns using a user-defined reference sequence to the user clustering to specified item or item clustering to specified user. The dissimilarity degree of the sequence of support values of an item set to the reference sequence is used to capture temporal prevalence variation matches the reference pattern of the user information or item information. The subset specification is used to define a user interest temporal pattern and guide the degree of approximate matching of prevalence values of associated item sets for it. The straight-forward approach is to divide the mining process into two separate phrases. The first phrase computes the
support values of all possible item sets at each time point and generates their support sequences. The second phrase compares the generated support time sequences with a given reference sequence and finds similar item sets. In this step, a multidimensional access method such as an R-tree family can be used for a fast sequence search. By exploiting interesting properties such as an envelope of support time sequence and a lower bounding distance for early pruning candidate item sets, an algorithm for effectively mining similarity-profiled temporal association patterns is modelled in this work which is correct and complete in the mining results and in providing the computational analysis.

9. **Vojnovic M, Cruise J, Gunawardena D, Marbach P (2009), “Ranking and Suggesting Popular Items”**. In this literature, ranking the popularity of items and suggesting popular items based on user feedback is considered. The true preference refers to the preference over items that would be observed from the users’ selections over items without exposure to any suggestions. A simple scheme for ranking and suggesting popular items analysis that suggests such a simple scheme can lock down to a set of items that are not the true most popular items if the popularity bias is sufficiently large, and may obscure learning the true preference over items. Alternative algorithms are designed to avoid such reinforcements and provide formal performance analysis of the ranking limit points and popularity of the suggested items. User feedback is obtained by iteratively presenting a set of suggested items, and user selecting items based on their own preferences either from this suggestion set or from the set of all possible items. The goal is to quickly learn the true popularity ranking of items (unbiased by the made suggestions), and suggest true popular items.
The difficulty is that making suggestions to users can reinforce popularity of some items and distort the resulting item ranking.

10. **Wickramaratna, K. Kubat, Miroslav Premaratne, (2009), “K.Predicting Missing Items in Shopping Carts”**. In this paper, precision prediction of unknown items can play a very important role which has focused mainly to expedite the search for frequently co-occurring groups of items in multiaspects filtering type of transactions for prediction purposes. It is predicted based on the presence or absence of other items in the transaction dataset. It is important to understand that allowing any item to be treated as a class label presents serious challenges as compared with the case of just a single class label. The number of different items can be very high, perhaps hundreds, or thousand, or even more. To generate association rules for each of them separately would give rise to great many rules with two obvious consequences: first, the memory space occupied by these rules can be many times larger than the original database (because of the task’s combinatorial nature); second, identifying the most relevant rules and combining their sometimes conflicting predictions may easily incur prohibitive computational costs. This paper contributes to the task of predicting the rating of the similar item with partial information about the contents of an other product with similar rating using decision support system or association rule Mining in a computationally efficient manner, all rules whose antecedents contain at least one item from the incomplete similarity. Then these rules are processed to estimate uncertainty of processing techniques, including the classical Bayesian decision theory based on the Dempster-Shafer (DS) theory of evidence combination.
11. Song Jie Gong (2010), “A Collaborative Filtering Recommendation Algorithm Based on User Clustering and Item Clustering”. In this paper, personalized recommendation systems can help people to find interesting items and they are widely used with the development of textile and movie industry. Many recommendation systems employ the collaborative filtering technology, which has been proved to be one of the most successful techniques in recommender systems in recent years. With the gradual increase of customers and products in electronic commerce systems, the time consuming nearest neighbor collaborative filtering search of the target customer in the total customer space resulted in the failure of ensuring the real time requirement of recommender system. At the same time, it suffers from its poor quality when the number of the records in the user database increases. Sparsity of source data set is the major reason causing the poor quality work. Utilizing the system to avoid the Sparsity and scalability in collaborative filtering, personalized recommendation approach is analysed to utilize the item clustering and user clustering as collaborative filtering mechanism to produce the recommendations’ Users are clustered based on users’ ratings on items, and each users cluster has a cluster center. Based on the similarity between target user and cluster centers, the nearest neighbors of target user can be found and smooth the prediction where necessary. Then, the proposed approach utilizes the item clustering collaborative filtering to produce the recommendations. The recommendation joining user clustering and item clustering collaborative filtering is more scalable and more accurate. This literature is suitable for item similarity formulation and user similarity formulation.
12. **Yehuda Koren, Joseph Sill (2011), “Collaborative Filtering on Ordinal User Feedback”**. In this paper, collaborative filtering (CF) recommendation framework is based on viewing user feedback on products as ordinal, rather than the more common numerical view. Such an ordinal view frequently provides a more natural reflection of the user intention when providing qualitative ratings, allowing users to have different internal scoring scales. Moreover, assigning numerical scores to different types of user feedback would not be easy. The framework can wrap most collaborative filtering algorithms, enabling algorithms previously designed for numerical values to handle ordinal values. This framework is demonstrated by wrapping a leading matrix factorization CF method. The work utilize the system to predict a full probability distribution of the expected item ratings, a frame work is created on the item, wrapping a leading matrix factorization CF method is used to estimate the confidential level of prediction. This literature is suitable for item similarity and user similarity formulation and formulation based on the other user rating.

13. **Bartolini I, Zhenjie Zhang, Papadias, D (2011), “Collaborative Filtering with Personalized Skylines”**. In this paper, Collaborative filtering (CF) systems exploit previous ratings and similarity in user behavior to recommend the top-k objects/records which are potentially most interesting to the user assuming a single score per object where the object can be an item. Systems usually take two steps: 1) retrieve users who have similar rating patterns with another user; and 2) utilize their scores to return the top-k records that are potentially most interesting to another user. Conventional CF assumes a single score per object. However, multiple attributes induce the need to distinguish the concepts of
scoring patterns and selection criteria maybe rated in which case simply returning the ones with the highest overall scores fails to capture the individual attribute characteristics and to accommodate different selection criteria. A typical CF system cannot differentiate between the two users, and based on their identical scoring patterns would likely make the same recommendations to both. To overcome this problem, the system could ask each user for an explicit preference function that weighs all attributes according to her/his choice criteria, and produces a single score. In order to enhance the flexibility of CF, Collaborative Filtering Skyline (CFS), a general framework that combines the advantages of CF with those of the skyline operator is analysed as proposed system. CFS generates a personalized skyline for each user based on scores of other users with similar behavior. The personalized skyline includes objects that are good on certain aspects, and eliminates the ones that are not interesting on any attribute combination. Although the integration of skylines and CF has several attractive properties, it also involves rather expensive computations, the challenge through a comprehensive set of algorithms and optimizations that reduce the cost of generating personalized skylines. In addition to exact skyline processing, an approximate method that provides error guarantees is incorporated. Finally, top-k personalized skyline is utilized, where the user specifies the required output cardinality.

14. Hui-Feng Sun, Gang Yu, Guang Chen (2012), “JacUOD: A New Similarity Measurement for Collaborative Filtering”. In this paper, a similarity measurement for memory based collaborative filtering, each user or item is considered as a vector, similarity measurement in different multidimensional
vector spaces are constructed for it. Based on Euclidean distance, considering characteristics of similarity measurement in different multidimensional vector spaces Jaccard Uniform Operator Distance (JacUOD) is analysed to investigate the characteristics of similarity measurement for different multidimensional vector spaces using unified similarity to measure the similarity between items and users. This leads to better prediction accuracy and it uses (SCODVS) aimed for unified similarity. This paper is suitable for item similarity formulation and user similarity formulation with high prediction accuracy.

15. Hornick and Tamayo (2012), “Extending Recommender Systems for Disjoint User/Item Sets: The Conference Recommendation Problem”. In this paper, traditional recommendation engine data sets are characterized by users U and items I, where users explicitly rate some subset of I producing a sparse matrix R of ratings. The goal is to provide ratings for items in I that have not yet been rated by users in U. In contrast, the type of data is encountered in the conference recommendation problem Conjoint of User and Item decomposition technique that is an extension of preference-based recommender systems to recommend items from a new disjoint set to users from a new disjoint set is analysed. The assumption being that preferences exhibited by users with known usage behavior which can be abstracted by projections of user and item matrices, will be similar to ones of new (different) users where the basic environment and item domain are the same. System does not require no item ratings, but operates on observed user behavior. The Proposed paradigm consists of projections of both user and item data, and the learning of relationships in projected space. Once established, the relationships enable
predicting new relationships and provide associated recommendations. In the more general case, the recommendation models that 1) general and represent relationships at a higher, more abstract level, and 2) do not rely on having specific histories for actual items or users, but can leverage past instances representing the acquisition, rating, or attendance of other items by other users.

16. Lin Guo, Qinke Peng (2013), “A Combinative Similarity Computing Measure for Collaborative Filtering”. In this paper, The recommendation system is used to discover favourite items. This research is made to recommend the qualities using three similarity methods. (Cosine similarity, adjusted cosine similarity and Pearson correlation similarity) the proposed method, a combinative similarity measure considers the account of items that user co-rated. The stability degree to improve the accuracy of collaborative filtering both based on item is incorporated with the similarity metrics. This literature is suitable for item similarity formulation and user similarity formulation methods show its satisfactory performance with less computation complexity.

17. Chua F.C.T, Lauw H.W, Ee-Peng Lim (2013), “Generative Models for Item Adoptions Using Social Correlation”. This paper, discusses the issue in predicting the product from the decision of the other similar user with same characteristics carried out with help of opinion miming. In making adoption decisions, users rely not only on their own preferences, but also on friends or neighbours, hence system has incorporated with neighbourhood based CF or feedback based CF. The modeling social correlation on users item adoptions using user-user social graph and an item-user adoption graph, item similarity, user similarity can be achieved. The system is correlated with the friend user to
predict the item similarity using model item adaptation strategies using

Social correlation frame work considers a social
correlation matrix representing the degrees of correlation from every user to the
users friends in addition to a set of latent factors representing topics of interests
of individual users. Based on the framework, two generative models, namely
sequential and unified, and the corresponding parameter estimation approaches
have also modelled. From each model, the social correlation and hybrid methods
for predicting missing adoption is possible. One that reflects the importance of
each latent factor to users, and another that does the same for items. However,
this approach assumes that all items adopted by a user can be fully explained by
the user’s and items’ latent factors. Some users may primarily rely only on their
own latent factors in making adoptions. These users have high self-dependency.
However, most users rely on a mixture of self-dependency and social
correlation.

Filtering Algorithm Based on User Interest”. In this paper, the system can be
applied to various rating prediction using sparse data. To reduce the impact of
data growth, using user information an algorithm called collaborative filtering
(CF) based on user cluster is proposed which is to improve the user similarity
calculation method and extends the user item rating matrix. Algorithm could
describe the user similarity more accurately and alleviate the impact of data
sparseness in collaborative filtering algorithm. Then, allocate the target user to
the most similar cluster and generate its nearest neighbor set. At last,
recommend the top items most interested by the nearest neighbors to target users
based on their predicted rating for the items. This literature is suitable for item similarity formulation and user similarity formulation.

19. **Yi Cai, Ho-Fung Leung, Qing Li, Huaqing Min, Jie Tang, Juanzi Li (2014)** “Typicality-Based Collaborative Filtering Recommendation”. In this paper, the basic idea of user-based CF approach is used to find out a set of users with similar patterns to a given user (i.e., “neighbors” of the user) and recommend to the user those items that other users in the same set like, while the item-based CF approach aims to provide a user with the recommendation on an item based on the other items with high correlations (i.e., “neighbors” of the item). In all collaborative filtering methods, it is a significant step to find users’ (or items’) neighbors, that is, a set of similar users (or items). Currently, almost all CF methods measure users’ similarity (or items’ similarity) based on correlated items of users (or common users of items). A better way to select “neighbors” of users or items for collaborative filtering can facilitate better handling of the challenges through the idea of object typicality from cognitive psychology and analyse a Neighborhood based CF recommendation approach. The mechanism of typicality-based CF recommendation is as follows: First, cluster all items into several items groups. Second, form a user group corresponding to each item group (i.e., a set of users who like items of a particular item group), with all users having different typicality degrees in each of the user groups. Third, we build a user-typicality matrix and measure users’ similarities based on users’ typicality degrees in all user groups so as to select a set of “neighbors” of each user. Then, predict the unknown rating of a user on an item based on the ratings of the “neighbors” of at user on the item. A distinct feature of the typicality-
based CF recommendation is that it selects the “neighbors” of users by measuring users’ similarity based on user typicality degrees in user groups, which differentiates it from previous methods. To the best of our knowledge, there has been no prior work on using typicality with CF recommendation. Proposed framework which provides a new perspective to investigate CF recommendations has the following several advantages: . It generally improves the accuracy of predictions when compared with previous recommendation methods. . It works well even with sparse training data sets, especially in data sets with sparse ratings for each item. It can reduce the number of big-error predictions. It is more efficient than the compared methods.

20. **Tang Zhi-hang, Zhang Min-min, Ouyang Wen-min(2015), “Personalizing Recommender System Based on Neighbourhood Collaborative Filtering”**. Collaborative methods have been categorized as memory-based if they operate over the entire data to make predictions and as model-based if they use the data to build a model which is then used for predictions. Among various recommendation techniques, neighborhood- based Collaborative Filtering (CF) techniques have been one of the most widely used and best performing techniques. A new approach has been proposed to enhance the neighborhood-based CF techniques to identify the few best neighbours. The Personalized recommender systems are used to list the ordered data or the score of items, and the goal is to assist the user in the decision-making process. The decision-making process, recommender systems use the available data on the items themselves. Personalized recommender systems subsequently use this input data, and convert it to an output in the form of ordered lists or scores of items.
in which a user might be interested. These lists or scores are the final result the user will be presented with, and their goal is to assist the user in the decision-making process. The application of recommender systems outlined was just a small introduction to the possibilities of the extension. Recommender systems became essential in an information- and decision-overloaded world. "This literature is suitable for item similarity formulation and user similarity formulation."