Chapter 9

Conclusions and Future Work

Having done the exhaustive study of recommender systems belonging to various domains, stock market prediction systems, social resource recommender, tag recommender for social bookmarking system and hybrid movie recommender have been developed. These developments ensure the application of recommender system in various domains. Summary of work done, conclusions derived therein and possible future work are mentioned herewith in detail.

- Forecasting in stock markets is one of the domains focused in the thesis. Two forecasting systems are developed for prediction in stock markets. First system addresses the problem of predicting direction of movement of stock market index and stock price. The focus of the second forecasting system is on predicting future values of stock market index. Predictions from these forecasting systems can be used to provide a kind of financial service to the users. This may enable the users to make correct decisions while investing in stock markets.

It was identified from the literature, that, machine learning techniques can be used to develop prediction models for stock markets. Artificial Neural Network (ANN), Support Vector Machine, Random Forest and Naive Bayes classifier are used in this thesis to develop the prediction models. It is also felt that improvement in representing and preparing the data which is to be used as the input to the prediction models, can improve the prediction performance. This is achieved by means of Trend Deterministic Data Preparation Layer and Two Stage Fusion Models for the task of predicting direction of movement and value
of stock market index respectively. Experimental results are in alignment with the intuition behind which these Trend Deterministic Data Preparation Layer and two stage fusion models are proposed.

Improvement of accuracy with the help of Trend Deterministic Data Preparation Layer which is based on common investor’s methods for stock investing, promotes the idea of pre-processing the data based on the domain in which machine learning algorithms are used. This idea can be further extended not only in stock domain by incorporating other human approaches of investing but also in various other domains where recommender systems and machine learning techniques are used. Ten technical indicators are used to construct the knowledge base, however, other macro-economic variables like currency exchange rates, inflation, government policies, interest rates etc. that affect stock market can also be used as the inputs to the models or in construction of the knowledge base of a recommender system. Average volume of a stock is also a potential candidate that may be useful in deciding the trend. It is worth noticing that at Trend Deterministic Data Preparation Layer, technical indicators’ opinion about stock price movement is categorized as either ‘up’ or ‘down’. Multiple categories like ‘highly possible to go up’, ‘highly possible to go down’, ‘less possible to go up’, ‘less possible to go down’ and ‘neutral signal’ are worth exploring. This may give more accurate input to prediction algorithms. The other careful observation reveals that the focus while predicting movement is short term prediction. Long term prediction can also be thought as one of the future directions which may involve analysis of stock’s quarterly performance, revenue, profit returns, companies organizational stability etc. In this thesis, for predicting in stock market, technical indicators are derived based on the period of last 10 days (e.g. SMA, WMA, etc.). It is worth exploring the significance of the length of this period, particularly, when the objective is long term prediction.

In the work related to predicting future values of stock market indices, design parameters of SVRs in the first stage are determined experimentally, however it may be worth exploring algorithms such as genetic algorithm to tune the design parameters of these SVRs. This may lead to more accurate prediction
of statistical parameters by these SVRs. Another direction for future work can be to use more statistical parameters as inputs to find much better correlation.

- The problem of social resource recommendation under the situation, where only positive and unlabeled examples are available, is also addressed in this thesis. Social bookmarking sites such as Bibsonomy or Delicious allow users to bookmark URLs and submit the research articles. User bookmarking a resource (URL) or submitting a resource (research article) on this system, implicitly indicates his liking to the resource. Other resources (URLs/research articles), however, do not imply negative preference of the user about them. This leads to the situation where we have positive examples, but no negative examples for user preference.

The memory based collaborative filtering has served as the most widely used technique for resource recommendations, but it has its own limitations of reliance on ad hoc heuristic rules and dependence of success on availability of a critical mass of users. If a learning based approach is to be devised, as the alternative to memory based collaborative filtering for the task of social resource recommendation, it requires to learn the recommender from positive and unlabeled examples.

Hence, two step methods based on naive Bayes classifier and CART/LS-SVM to learn a recommender using positive and unlabeled examples are proposed. Moreover, a direct method in which LS-SVM is adapted to learn from positive and unlabeled examples is also proposed. Experimental results validate the theoretic assumption behind the proposal of this scheme and optimal results are achieved.

A recommender that can exploit content data of the resources to generate recommendation may be an interesting direction for the future work. Another direction for the future work is to fuse the recommendations from the usage based and content based recommender to improve the performance.

- It is known that, social bookmarking system allows users to upload, share and organize their resources. It also facilitates user to annotate his resource with free form tags. The freedom of selecting tags compels user to write descriptive tags
on his own to define his viewpoint, which is burdensome and time consuming task. Hence, user may avoid or assign very small number of tags to resource, resulting in very sparse tag space. Further, different users may choose tags based on their knowledge background and preferences i.e. they may describe the same resource based on different granularity level resulting into noisy tag space and create difficulty to find relevant material based on such tags. It is important to notice that synonymous tags increase data redundancy and polysemous tags i.e. a tag that has many contextual meanings, lead to inappropriate connections between resources.

The effective solution is Tag Recommendation System that automatically suggests appropriate set of tags to user while annotating resource. Hence, a tag recommender is implemented to assist the user in tagging process. The recommendation task is modelled as multi-label text classification problem. Textual content of the resources is used to learn the tag recommender. Using the textual content, resources are represented with various Information Retrieval models such as Boolean, bag-of-words and TFIDF. It is evident from the result that, the tag recommender performs the best when it is learnt through the bag-of-words representation of the resources.

A personalized tag recommender is definitely an interesting direction for future work.

- Movie recommendation is an open research area with unanswered problems and with growing social networking data. A movie recommender system, that is solely based on ratings or content of the movies may not be accurate enough. There is a need of systematically fusing different types of data about movies and users from various sources to improve the quality of recommendations.

Hence, an item-based hybrid filtering approach which combines usage, tag and content data of movies is proposed. Movie recommendation task is modelled as classification problem where it is predicted that the user will like or dislike the movie. Movie Recommender system proposed in the thesis exploits movie specific data such as movie genres, star cast and directors in addition to the ratings and weighted tags. Item profiles are constructed from the careful combinations
of different types of data. Similarity between items is then calculated based on these profiles. It is evident from the results that fusing the different kind of data related to movie, appropriately, improves the accuracy of the recommender.

A possible direction for the future work may be the use of machine learning techniques for exploiting these profiles and generating recommendations.

- In a recommender system, the items which are liked or disliked by the user are the labeled examples about the users’ preference. However, there are many other items which are not rated by the user. These items form the set of unlabeled examples. One of the major problems with recommender system is that, if it does not have adequate number of labeled examples for the user for whom recommendations are to be made, it may not be adequately accurate and useful.

The usefulness of unlabeled examples in learning the recommender in the presence of small number of labeled examples is crucial and hence, studied in this thesis. From the experimental results, it is evident that, accuracy of the recommender is improved when it is learnt using unlabeled data in addition to the labeled data.

A interesting issue is that researchers have yet not shown that when the labeled data set is sufficiently large, the unlabeled data still help. This can serve as the useful future direction. There is also a need to evaluate the usefulness of unlabeled examples in different domains and on different datasets in order to reach to some concrete conclusions.