



Survey on Kernel Optimization based Enhanced Preference Learning for Online Movie Recommendation

Sreelekshmi. B

M.Tech, CSE, Sree Buddha College of Engineering, Elavumthitta, Kerala, India

Vidya. N

Assistant Professor, CSE, Sree Buddha College of Engineering, Elavumthitta, Kerala, India

ABSTRACT

A social recommendation system has attracted a lot of attention recently in research communities. Traditional social recommendation algorithms are often based on batch machine learning methods that suffer from several critical limitations, for example, a cost of training is extremely expensive in each time new user ratings arrive, unable to capture the change of the user's preferences over time. The proposed system is a new online social movie recommendation framework from the view point of online graph regularized user preference learning (OGRPL), which incorporates the user-item collaborative relationship as well as content characteristics of elements in a unified process of learning preferences. This method overcomes the problem of excessive data adjustment by eliminating the low range constraint. However, it does not handle uncertain ratings that contain irregularities. The propagation of the Monte-Carlo uncertainty can be used to handle uncertain data in the preferences modeling. The level of uncertainty is measured. Also use SVM (Support Vector Machine) to predict user preferences using the model learning method. SVM is a classification method which is used to classify movies. Then based on classification recommendation is performed.

Keywords: Online social recommendation, user preference learning, low rank matrix, uncertainty, over fitting problem

I. INTRODUCTION

In today's digital world, there is an infinite variety of contents to consume such as books, videos, articles, movies, etc. Finding the content that peoples like has become a crucial task. This is where the recommendation system comes from. Content providers recommend contents to users based on user's interest. A recommendation system or recommender system is a subclass of the information filtering system that seeks to predict the "rating" or "preference" that a user would give to an item. Recommendation systems address the problem of information overload by filtering group of important information from a large amount of information generated dynamically according to the user's preferences, interest or behavior observed on the content. It is a system that can be used to create a user profile. The use of efficient and accurate recommendation techniques is very important and necessary 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. Recommender systems typically produce a list of recommendations in three ways:

1. Content-based Filtering
2. Collaborative Filtering
3. Hybrid Recommender systems

An approach in designing recommendation systems is filtering based on content i.e. content based filtering. Content is the basis for creating recommendation applications. Content-based filtering methods are

based on a description of the content and a profile of user preferences. In a content-based recommendation system, keywords are used to describe the elements and a user profile is constructed to indicate the type of item that this user likes. In other words, these algorithms try to recommend articles similar to those that a user liked in the past (or that he is currently examining). In particular, different candidate elements are compared with the elements previously rated by the user and the contents that best match are recommended. This approach has its roots in information retrieval and information filtering research.

Another common approach to the design of recommendation systems that has a wide use is collaborative filtering. Collaborative filtering methods are based on the collection and analysis of a large amount of information about the behaviors, activities or preferences of users and the prediction of what users like based on their similarity with other users. A key advantage of the collaborative filtering approach is that it does not depend on the analyzable content of the system and, therefore, is able to accurately recommend complex elements, such as movies, without requiring an "understanding" of the article itself. Many algorithms have been used to measure the similarity of the user or the similarity of the element in the recommendation systems.

Hybrid recommendation systems are adjusted to join collaborative filtering (CF) based on content that controls in a framework, and increases the benefits and also decreases the weaknesses of both techniques. Therefore, hybrid recommendation systems work on characteristics that are related to both. It can provide more precise recommendations than pure approaches. Hybrid approaches can be implemented in several ways: making content-based and collaborative-based predictions separately and then combining them; adding content-based capabilities to a collaborative approach (and vice versa); or by unifying the approaches into one model. Also be used to overcome some of the common problems in recommender systems such as cold start and the sparsity problem.

Kernel Optimization Based Online Movie Recommendation is a novel framework of movie recommender system termed Online Graph Regularized User Preference Learning (OGRPL). OGRPL is a hybrid model using both CF information via the partially observed user-item matrix as well as

the auxiliary content features for each item. Given a stream of user ratings, OGRPL incrementally learns the user preference on the content features of the items. However, human beings are prone to making error ratings and rating data always contain noise in practice. Therefore, direct learning of user preferences may be an over-fitting problem and therefore not robust. OGRPL-Frank Wolfe (OGRPL-FW) is used to solve this problem. As a result, get more accurate movie recommendations for users based on their preferences. Online graph for the learning of user preferences (OGRPL), which incorporates both the collaborative user-article relationship, and the characteristics of the content of the object in a unified preferences learning process. This method overcomes the problem of over-fitting by eliminating the low rank constraint. Adopt online learning to recommend movies. However it does not handle the uncertain data that contain irregularities in rating. This means that a big difference in ratings between different users will affect performance. Therefore, in preference modeling applications, the modeling functions can be formulated using the non-linear function. Monte-Carlo uncertainty propagation can be used to manage uncertain data in the modeling of preferences. The level of uncertainty is measured and considered for ranking preference [1].

Also SVM (Support Vector Machine) can be used to predict the user preference by model learning method. A training set may be created and kept for offline process. An SVM library (libSVM) is used to learn the test data using training data. The performance of SVM is greater than Frank-Wolfe algorithm. There are different kernel type that can be applied on SVM. The kernel selection generates different output and the best kernel is selected. Figure.1 shows architecture of the system.

II. LITERATURE SURVEY

Many recommendation systems have been developed in the past. These systems use different approaches like collaborative approach, content based approach, a utility base approach, hybrid approach etc. Most traditional recommendation algorithms are based on batch training techniques. In this techniques assume all user ratings are given in the form of user-item matrix. Such assumption unsuitable for real-world online recommendation applications. The user ratings arrive sequentially in an online applications. Whenever new ratings are arrived, making the

training process extremely time-consuming. Moreover, if the size of the data is too large, it is very difficult for handling all the data in the batch mode. Also user preference could change over time in real world online application, which makes the batch learning processes fail to capture such changes dynamically. Some of the movie recommendation systems are:

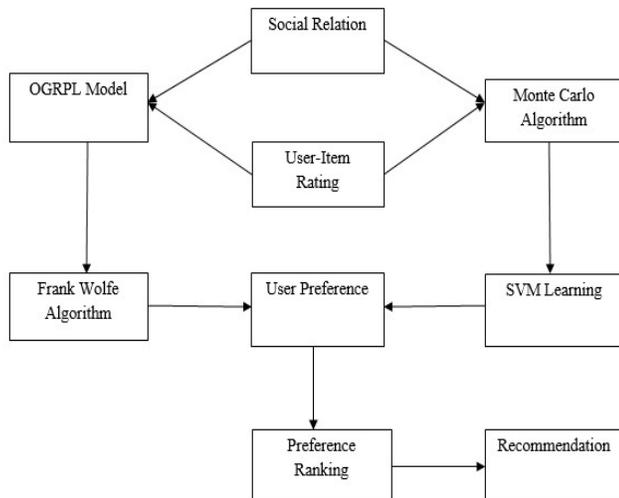


Figure.1: Architecture of the system.

P. Li et al. [2] proposed a new framework of movie recommendation based on content based filtering. This recommendation system is based on genre description or typeabout movies rather than similarity of other users. Recommended movies are similar to users who rated, visited, browsed and liked movies in the past. In movie recommendations, the number of feature factors, such as actors, directors, stories, song sequences, etc., may be considered, which may be used to recommend movies based on user interests. Assume that two feature factors, "romantic" and "action", classify the movie. Comparing the user profile's interest and feature factors, the movie can be recommended to the user. For example: User X prefers romantic movies more than action movies. X has rated Movie2, Movie3 as 5 and Movie1, Movie5 as 0. Movie4 can be recommended to X by comparing feature values, because it is tagged as a romantic movie. The main limitations of content-based filtering are limited content analysis, over-specialization, and cold-start issues. Limited content analysis makes it difficult to analyze a set of features that are representative of the complete content. If they contain the same terms and phrases, it is very difficult to analyze whether it is good or bad. It does not guarantee the quality of the content. Over-specialization means that the same event can be called a different headline. In cold start problem, items are

not recommended because of lack of information about users and contents.

Manoj Kumar et al. [3] suggested a movie recommendation system named MOVREC. It is based on a collaborative filtering method that uses user-supplied information, analyses them, and then recommends the best movie for the user at the time. The list of recommended movies is sorted by the previous user's ratings for these movies and uses the K-means algorithm for this purpose. Due to the simplicity, flexibility and computational efficiency of the K-means algorithm, especially considering a large amount of data, it is one of the most commonly used clustering methods. K-means repeatedly calculates k cluster centers and assigns objects to the nearest cluster based on distance metrics. When there is no more change in the center point, the clustering algorithm will converge. However, K-means lacks the ability to select suitable initial seeds and may result in inaccurate classification. Randomly select the initial seed it may cause the local optimal solution to be worse than the global optimal solution. MOVREC can also help users find movies of their choice based on the movie experience of other users in an efficient and effective way, without wasting a lot of time for useless browsing. The main contribution of the system is that the proposed system has been tested by a small group of people and the system has received their positive response. Keep the system simple and interactive. In order to recommend the movie to the user accurately, a K-means clustering algorithm and a pre-filter are applied. To give weight to attributes and give them priority, they surveyed a group of people and prioritized attributes based on the results obtained. However, this method has some problems due to the use of collaborative filtering such as gray sheep, data sparsity, and early rater issues. K-means also has premature convergence.

D. Pathak et al. [4] suggested a movie recommendation system called ORBIT. Today, users will get so many movie recommendation websites to recommend best movies to users based on their interests. All of these sites implement one of the traditional content, context, and collaboration recommendation algorithms. Alone, these algorithms fail to recommend the best and effective recommendations to the user. Therefore, it is necessary to develop a unique algorithm that combines the features of traditional algorithms with their new features. ORBIT is a movie recommendation system based on a unique hybrid

recommendation algorithm that satisfies user's needs by providing optimal and effective recommendations. It also contains context-based input. This type of input specifies time or behavior-oriented data, such as date, weather, taste, mood, etc. In general, context-based recommendations are in such inputs. It provides an easy-to-use graphical user interface for user profile and movie information management. It provides the best recommendations for those who do not have sufficient personal experience or ability to assess the number of alternatives offered by the site. The main drawback of this system is that it is very difficult to find relevant information from a large amount of available data, the computational complexity is high and the implementation cost is high.

Costin-Gabriel Chiru et al. [5] introduced a movie recommender system which provides movie recommendations based on the user's psychological profile. These recommendations are done by analyzing the user's psychological status, their viewing history, and movie scores from other websites. They are based on aggregate similarity calculations. The system uses both collaborative filtering and content filtering (using methods based on the different features of the movie in the database). Although similar applications are available, they tend to ignore user-specific data, which is essential for his/her behavior. The app named movie recommender is a hybrid recommendation system that provides movie recommendations by analyzing the user's psychological state. It is due to the observation of the lack of such a recommendation system on the Internet. Its purpose is to provide recommendation specific to the user and to observe whether there is a real connection between a person's psychological condition and his/her preference for the movie. In content-based filtering, psychological features are used for one of the content-based features and are determined with the help of yes/no psychological tests. The purpose of this test is to determine to what extent psychological profiles can be found (choleric, sanguine, phlegmatic, and melancholic) the user's own personality. The psychological profile is choleric, characterized by its outgoing nature, hot-tempered and quick thinking, active, practical, strong-willed, self-confident, independent, and recommended movies for this feature are horror, war, action, comedy and romance. Optimism, impulsiveness, fun, persuasiveness, lightheartedness, and acceptance are the characteristics of excitement. It is characterized by its extroverted characteristics. The films that

recommend this feature are action, adventure, comedy, drama, and romance. It features introverted, calm, ruthless, easy to use, slow, indirect, practical, patience, perseverance and consistent movie, recommended to this archive are science fiction, tragedy, cult, drama and film. Collaborative filtering, as part of the collaborative approach, determines the similarity between using the database and the user's movie, and calculates based on the history of movie ratings. The similarity between movies is calculated using the cosine similarity between each movie the user had previously seen and other movies from the database.

When the recommendation system is implemented, the user's preference for the movie may also be very relevant, so the last similarity value to be evaluated is the user similarity determined using the nearest neighbor algorithm. Using this method, users with similar preferences to the current user are found based on their psychological status. After that, the user closest to the current user is determined and the movie that they rated as positive and the current user cannot see is extracted. These will become candidates recommended to the current user. Obviously, the movie that the closest user rated as positive will represent a more reliable recommendation for the current user, so their value will be closer to 1 than the value of other movies. However, this method requires less film features, insufficient time to collect enough information to reliably predict user preferences, and it is not possible to reliably assess the accuracy of the system.

Shinhyun Ahn et al. [6] developed a simple and low-cost movie recommendation system harnessing vast cultural metadata, about movies, on the Web. Cultural metadata is defined as "information implied by massive amounts of data that need to be extracted using information retrieval techniques". This system uses five kinds of cultural metadata - user reviews, plot summary, episode summary, plot keywords and genre. User reviews are people writing reviews about the movies they see and sharing their impressions and opinions with others. User reviews are indeed rich cultural metadata. Drawing outlines are user-created brief summaries of 5-10 lines, rather than judging the quality of the movie, because it directs that the outline of the episode should not include any analysis of the movie. Therefore, it is expected that the plot summary only represents the plot of the movie. However, despite its huge potential, drawing outlines are still

sparse metadata. The storyline is longer than the plot outline, but it is less scary than the plot outline. The story key is a controlled keyword related to the movie story. The plot keywords are managed and controlled by IMDb and therefore have reliable cultural metadata. The genre is also the control metadata provided by the IMDb. Each movie is based on several genre categories represented by "Action / Adventure / Family / Fantasy / Science Fiction".

This five metadata are divided into two types, text-type and keyword-type. Text-type metadata such as user comments, plot outlines, and synopses for movies are collected by web crawling. Then analyzed text-type metadata morphologically. At first it was needed to remove some 'language noises' to create document vectors for each text-type metadata which are natural languages. Then regulated words of metadata by Part-Of-Speech tagging and lemmatizing. Plural nouns were changed into singularities. Then words out of interest were filtered out. Chose nouns or modified words, or both. This is based on empirical common sense that nouns and modified words represent best the impressions of movies. Filtered words form a term pool for each movie, which represents the attributes of the Movie. From a term pool, created a document vector for each movie by using a TFIDF (Term Frequency Inverse Document Frequency). Mood vectors is a common sense reasoning toolkit and provides a vast common sense DB And several reasoning functions, one of which is a 'Guess Mood' function that outputs a mood vector for input text. Each mood vector has 6 components, (angry, surprised, disgusted, happy, sad, fearful), which represent the degree of emotions. Keyword-type metadata including plot keywords and genres are controlled vocabularies, therefore keyword vectors were created through TFIDF term weighting without an additional regulation. Computed similarities between movies with the above 5 types of metadata to make a recommendation list.

Chen et al. [7] introduced movie recommendation based on artificial immune system. The system combines an artificial immune system with collaborative filtering based on users and items to create an efficient, accurate recommendation system. By using the characteristics of antibodies and antigens in the artificial immune system (AIS), Kendall Tau and Weighted Kappa two different affinity measurement algorithms to calculate the correlation coefficient of the movie recommendation system.

Collaborative filtering then effectively finds useful users and items for scoring predictions. AIS, which has proven to be an effective clustering and classification algorithm, can be well applied to collaborative filtering and achieve better performance in recommendation. Thus, artificial immune system algorithms are incorporated into collaborative filtering to find clusters of similar users or similar items. Through this algorithm, the efficiency and accuracy of prediction in the recommendation system can be improved. The artificial immune system algorithm stems from the mechanism by which the biological immune system resists foreign pathogens. By using an adaptive immune response, this algorithm can be used to search for the best solution to the optimization problem. It can also be used to cluster or classify data to solve problems. In the real immune system, when a biological organism encounters an invasion from an external pathogen, the biological organism develops its own protective agent called an antigen. The immune system produces antibodies that bind to antigens to protect against pathogenic immune responses. This process of the immune system provides a good way to find the best solution to solve the optimization problem. Antigen is an objective function and antibodies are the solution to this problem. Two movie recommendation systems use collaborative filtering techniques and AIS implementations, one using a weighted Kappa method to calculate correlation coefficients, and the other using the Kendall Tau method. The AIS was established in order to select a group of people with similar film preferences as a goal. People in the database are considered candidate antibodies, and users using the movie recommendation system are considered antigens. Two related methods are used to calculate the correlation between antigens and antibodies, and the correlation between antibodies and antibodies. It consists of the following steps. There are some people's preferences stored in the database. The user enters his preferences for the movie and needs recommendations for some movies he has not seen. AIS selects a group of people who have similar preferences with the user. The weighted average of the group's preferences is calculated by the CF to generate the user's desired recommendations.

This method uses two different algorithms to calculate affinity (correlation coefficients) - Kappa and Kendall heads. The Weighted Kappa algorithm is a method of calculating affinity (correlation coefficients), which is calculated based on observed and expected

frequencies. The Kendall head is another method for calculating affinity (correlation coefficient). After the AIS selects a person (antibody) with similar preferences to the user (antigen) needs to be recommended, in the recommendation generation phase, CF will be used to calculate the prediction. The concentration is used as a weight to calculate the prediction because it contains the correlation of the antibody with the antigen and the correlation of the antibody with other antibodies. This method uses a Weighted Kappa method to obtain antibodies for one user (antigen) and uses the Kendall tau method to calculate the correlation between antibodies and antigens. The conclusion of this method is that Kendall's head is not suitable for the movie issue, and Kappa is a good method for movie problems. However, the main problem with this method is the low prediction accuracy.

Wei Yang et al. [8] proposed a movie recommendation system in heterogeneous network. For large movie data sets and various user interests, the recommendation system should pay attention to time consumption and personal interests. However, the existing information filtering technology rarely studies the user's interest in the movie. In order to effectively implement personalized movie recommendations, user interest-based movie recommendations are proposed in the heterogeneous network (IMRHN). It proposes the mutual influence of IMRHN with user's information and users in order to achieve personalized recommendation. This system mines the user's potential interest in movie types based on heterogeneous networks. Then divide the idea to reduce the time spent processing the actual data set and ensure proper accuracy. First, the heterogeneous network made up of movies, movie types and users is divided into multiple sub-networks based on movie types, and then the interest of active users in each movie type is explored through the influence of other users. Finally, a higher marked movie is recommended for each eligible movie type to attract the interest of more active users. Mining out the interest rate of the active user for the movie type. Based on the movie's average rating, the system will recommend the first m movie pair active users belonging to the selected movie type for each selected movie type system. However, this method requires more time to process large amounts of data, and the computational complexity is very high. The recommendation system applies intelligent access technology to large information systems. Systems,

especially those based on collaborative filtering, are gaining widespread success on the Web. In recent years, the amount of information available on the site and the number of visitors are increasing dramatically. Even very large-scale information resources require new recommendation system technologies to quickly generate high-quality recommendations.

Peng Li et al. [9] introduced a novel recommendation system based on inductive learning. Rather than calculating the similarity of user users or contents, the system builds decision trees to represent user preferences. It is recommended to perform the classification through the decision tree. The C4.5 algorithm is used to construct a decision tree. In this system, a recommendation technology based on induction learning is proposed to try to solve the problems of sparsity, scalability and transparency. The basic idea can be described in three steps. One is to correlate the evaluation value of the project and its attributes with the categories and attributes of the decision tree. Inductive learning is then performed to build a decision tree that represents the user's preferences. And through the decision tree classification to obtain the predicted value of the new item. Recommendation techniques based on inductive learning solve sparse problems by sharing the properties of each item. The computational cost in the inductive learning process is sufficiently low that a decision tree can be constructed immediately. In addition, the calculation cost in the recommendation process is linearly related to the total amount of available items. So, as scalability expands, it won't become huge.

Zan Wang et al. [10] constructed an optimized clustering algorithm for movie recommendation system using computational intelligence to partition user profiles which have been represented by denser profile vectors after Principal Component Analysis (PCA) transforming. The entire system consists of two phases, namely the online phase and the offline phase. In the offline phase, the clustering model is trained in a relatively low dimensional space and is ready to locate active users in different clusters. In the online phase, a TOP-N movie recommendation list is presented for active users due to the predicted rating of the movie. In addition, a genetic algorithm (GA) was used in the new method of improving the performance of the K-means clustering. The improved clustering algorithm was named GA-KM. Genetic algorithms are inspired by the theory of natural

evolution, which is known for its global adaptability and powerful search capabilities to capture good solutions. Due to its random search in large and complex spaces, it can efficiently solve various optimization problems. Darwin's natural survival principles guide the entire process of genetic algorithms and provide mechanisms for simulating biological evolution. Genetic algorithms use a group of "individuals" as chromosomes to represent possible solutions to a given problem. Each chromosome contains a number of genes that are used to calculate fitness to determine the likelihood of the next generation of reproduction. However, in the K-means algorithm, in the K-means algorithm, the initial seed's sensitivity choice affects the final output.

Rupali Hande et al. [11] introduced a movie recommender system MovieMender. MovieMender's goal is to provide users with accurate movie recommendations. In general, the basic recommendation system considers one of the following factors to generate a recommendation; user preference (ie, content-based filtering) or similar user preference (ie, collaborative filtering). In order to establish a stable and accurate recommendation system, a mixture of content-based filtering and collaborative filtering will be used. Due to the various shortcomings of pure content and a pure CF-based system, the system is a hybrid recommendation system, which is referred to as a content-enhanced collaborative filtering system. The hybrid system uses the representation of the content and the similarities between users. The intuition behind this technique is to use content-based predictors to populate sparsely distributed user rating matrices. A web crawler is used to download the necessary movie content for the data set. After preprocessing, the movie content database is stored. The data set consists of a user rating matrix. Content-based prediction is used to train each user rating vector in the user rating matrix and convert it to a pseudo-rating matrix, combining the actual rating with the prediction rating. Then apply collaborative filtering to this complete pseudo-user rating matrix to provide recommendation for active users.

To overcome all the difficulties of the existing methods, develop a novel framework of social movie recommender system termed Online Graph Regularized User Preference Learning (OGRPL). It is a hybrid model combining both collaborative and content based filtering to recommend movie to users based on their preference. The contributions of this

system are Present a new framework of online movie recommendation from the viewpoint of graph regularized user preference learning, which incorporates both collaborative user-item relationship as well as item content features into a unified preference learning process. Develop an efficient iterative procedure, OGRPL-FW which utilizes the Frank-Wolfe algorithm, to solve the proposed online optimization problem.

CONCLUSION

A social recommendation system has attracted a lot of attention recently in the research communities. The proposed system is a new framework of online social movie recommendation from the viewpoint of online graph regularized user preference learning (OGRPL), which incorporates both collaborative user-item relationship as well as item content features into a unified preference learning process. This method overcome the over fitting problem of the data by eliminating low rank constraint [1]. Monte-Carlo uncertainty propagation can be used to handle uncertain data in preference modelling. Uncertainty means irregularities or non-linearity in rating value. The level of uncertainty is measured. Also SVM (Support Vector Machine) can be used to predict the user preference by model learning method. A training set may be created and kept for offline process. An SVM library (libSVM) is used to learn the test data using training data.

REFERENCES

1. Zhou Zhao, Hanqing Lu, Deng Cai, Xiaofei He, and Yueting Zhuang, "Kernel Optimization based Enhanced Preference Learning for Social Recommendation," IEEE Trans. Knowl. Data Eng., vol. 28, NO. 9, Sep. 2016.
2. P. Li, and S. Yamada, "A Movie Recommender System Based on content based filtering," IEEE Conf. on Cybernetics and Intelligent System, pp.318-323, 2004.
3. Manoj Kumar, D.Yadav, Ankur Singh, Vijay Kr. Gupta," A Movie Recommender System: MOVREC" International Journal of Computer Applications. Vol.124 –No.3, August 2015.
4. D. Pathak, S. Matharia, and C. N. S. Murthy, "ORBIT: Hybrid movie recommendation engine," Emerging Trends in Computing, Communication and Nanotechnology (ICE-CCN),

2013 International Conference on, 2013, pp. 19-24.

5. Costin-Gabriel Chiru, Vladimir-Nicolae Dinu , Ctina Preda, Matei Macri, "Movie Recommender System Using the User's Psychological Profile," in IEEE International Conference on ICCP, 2015.
6. Shinhyun Ahn, and Chung-Kon Shi, "Exploring Movie Recommendation System Using Cultural Metadata", International Conference on Cyberworlds, 2008.
7. Chen, Qi and Aickelin, and Uwe, "Movie Recommendation Systems Using An Artificial Immune System," In: 6th International Conference in Adaptive Computing in Design and Manufacture, 2004, Bristol, UK.
8. Wei Yang, Xiaohui Cui, Jin Liu , Zhibo Wang, Weiping Zhu, and Li Wei, "User's Interests-based Movie Recommendation in Heterogeneous Network," IEEE Conf. on Identification, Information, and Knowledge in the Internet of Things, DOI 10.1109/IKI.2015.23.
9. Peng Li, Seiji Yamada "A Movie Recommender System Based On Inductive Learning." IEEE Conf. on Cybernetics and Intelligent Systems, Singapore, 1-3 December, 2004.
10. Zan Wang, Xue Yu, Nan Feng, and Zhenhua Wang, "An Improved Collaborative Movie Recommendation System using Computational Intelligence," IEEE Conf. on Computational Intelligence for Modelling, 0-7695-2504-0/05.
11. Rupali Hande, Ajinkya Gutti, Kevin Shah, Jeet Gandhi, and Vrushal Kamtikar, "MOVIEMENDER- A Movie Recommender System", International Journal of Engineering Sciences & Research Technology. Vol.11 – No.5, Nov. 2016.