Semester Long Project

Prithivirajan Karthikeyan(B00817066)

1) The solution to the recommendation system in this project works based on collaborative filtering using cosine similarity. Initially the given data is imported and the average rating of all the users are calculated. Then the ratings given by different users to different items are subtracted from the average rating obtained for different users. Then a pivot table is created with user ID as index, item ID as column and subtracted ratings as values. Then cosine similarity of the pivot table is calculated, and the entire table is subtracted from 1. Then most similar users to a particular user is obtained by collecting users with highest similarity score. Then the ratings of the highest similar users given to the item to be rated is got. Then these ratings are multiplied with the respective similarity score of the top similar user. Then these multiplied ratings and similarity scores are added. The value returned is divided by the addition of the similarity score of the top 5 similar users. By repeating this process to all the users and items the full dataset with all users and items and their respective rating can be obtained.

2) The solution to cold start issue in my project is that the calculated average rating value of a user is given in the place of rating for a particular user rating to a particular item whose value is missing.

3)

Example 1: Motion planning with probabilistic roadmaps

Matrix completion is one of the most important technique in probabilistic roadmaps (PRMs). Robots are automated vehicle can achieve moving from one place to another using probabilistic roadmaps which is based on matrix completion. Given a matrix, the system will draw multiple lines across the matrix and will predict which among the available route is more preferred and efficient.

Example 2: Object detection and classification.

Matrix completion plays a major role in object detection and classification. In object detection, a matrix is framed and from that the picture is divided based on the size of the matrix. Consider detecting a red colored ball. First a matrix is laid on the picture and from the grid which has the highest color coding of red and the ball is detected. Also, in classification, consider MNIST digit classification.

4)

Paper 1: Multi-Clustering Applied to Collaborative Recommender Systems

This paper discusses about applying multi clustering algorithm to a recommender system based on collaborative filtering. Different clustering algorithms will provide different views of analyzing the same data and the output can be obtained based on what is needed. As the content on the internet is getting wider day by day, it's a difficult task for the recommendation system to narrow down and predict what the particular user likes. The method described in this paper combines both content based and collaborative filtering approach to recommendation. There is a total of two process in this recommendation. First step prepares a step of similar neighbors using k – means clustering. This clustering is performed many times using the same k value. The results of this process are stored and is used as the input for the next process i.e., the recommendation process.

In the second step, the generating recommendation is performed. It starts from matching an active user with the best cluster as its neighborhood. Then similarity is calculated among the active user with the multiple users from each cluster. Then similarity measure is estimated in the same way like in the recommendation process. Then candidate items are searched by collaborative filtering item-based technique, but only within the cluster of neighborhoods.

Paper 2: Content-based recommender system for online stores using expert system

In this paper apart from just looking into the user rating other features are adding for recommending so that the recommendation will be more precise and accurate. Initially in this system when a new user is created the following information such as personal information (name and surname), login information, birth date, favorite categories of products are given to the system. Based on the given basic information the system will be able to give an initial prediction of what the new user will prefer to see. This is done by determining the user's similarity with other users by using any of the similarity determination techniques such as Pearson correlation coefficient. As a result, a set of users are obtained. From the similar users we got, the set of users with the same date of birth is selected. Next from the set of users the vector of favorite categories is specified. Then users with lesser similarity is excluded. Then the items viewed by the set of users are selected by categories that are of interest to the current active user. These items are sorted according to the frequency of their occurrence and are recommended to the current active user. After this initial recommendation, the system collects all the rating given by the user to different items. Apart from just rating this system also collects the other input data and gives out one output data. Input linguistic variables are: INP1 - the viewing time of the items, INP2 – how many times an item has been viewed by the user, INP3 - how many items of the same category the user was viewing. Output linguistic variable is: OUT – the popularity of an item. Thus this system makes recommendation more precise by collecting some additional data.

Paper 3: Surprise and Curiosity in A Recommender System

This paper aims in maintaining user engagement and sustain such engagement in the long run by using two human-centered objectives such as surprise and curiosity. This is implemented in a recommender system via two approaches. Knowledge-Based approach and Adaptive Knowledge-Based approach. Surprise in recommender systems aims to deliver richer information that is outside users expectation and curiosity is defined as a strong desire to explore, investigate, or learn something.

1) Knowledge-Based Approach: In KB approach, the expectation of seeing an article in the corpus is modeled by Pointwise Mutual Information (PMI). Let p(ti) and p(tj) are the probabilities that topics

ti and tj occurs in an article and PMI(ti,tj) = log2 [p(ti,tj)/[p(ti) * p(tj)]]. If the value of PMI is higher then it is more likely that both topics ti and tj co-occur within a single article in the corpus, and therefore a lower surprise score and vice versa.

2) Adaptive Knowledge-Based Approach: Suppose we have a recommended article ak based on the user's ui selected topic tij in a session sk. Also, suppose that ak is labeled by the set of topics Sak . If ak do not receive ui's interest, then further recommendations of articles labeled with topics in Sak based on KB method may result in losing user's attention. AKB method is like KB method but it incorporates user's ratings as the implicit real-time feedback. AKB uses the same selection approach, but it removes topics that are not interested in ui during previous sessions.

5) The major issue involved in this project is the whole project because the recommendation system is completely new, and I have to go through every aspect of what is involved in the recommendation system. But once I got into this the only issue was cold start. Apart from this project when I went through multiple articles there were many issues related with the process in implementing recommendation system. The major challenge next to cold start is synonym. It is a problem when the same item is given two or more names. In this situation the working efficiency of the recommender system will be reduced. For example, consider certain items are grouped under kitchen ware and certain items under kitchen utensils. But both are the same. In such a case the system will be confused, and it might not recommend certain items to users who want to buy it. The next issue face by the recommendation system is privacy. Recommender system is fully based on the information given by the users. So, it is a big deal for the organizations to maintain the user information. The most important issue face is shilling attacks. This issue is mainly when a person creates an account and gives false rating to the item. And just because the person information is similar to the user who gave false rating a list of wrong recommendation will be given to that similar user. Scalability is one of the issues faced by the recommendation system. Many sites such as amazon has a large amount of data which is really a big task to process. So, this leads to scalability issue. But this issue can be solved by reducing the dimensionality of the data collected.

Also considering in general the major problem faced by the online shopping sites such as amazon and social medias such as Facebook and Twitter are that the change in trends. Let us consider a situation when a new phone release. In this technological world most of the people will surely check on that phone and other related products. This makes the website like amazon think that this person is more into electronics and suggests more electronic based items. Nut that person likes to buy more of dresses and cosmetics. But once the person looks for phone or even buys it the recommendation system will recommend more similar phones to a person who likes dresses and cosmetics more. This is also the same case with a person who is gifting a friend who has different likings but just because this person buys it will be commended more. Apart from this even a single person might not be interested in the same item for a long time as his interest might change over time.