Midterm Project Report
Machine Learning with Networks: ECEN765
Recommendation System Based on Collaborative Filtering

Introduction & motivation:

Over the last decade, there has a rapid growth in the internet users. The growth in internet users between 2000-2018 is more than 1000 percent and it' still growing. Companies like Amazon, Flipkart, eBay, Netflix and YouTube introduce new stuff every single day.

There is a high probability of the user being deluged with the content getting uploaded into these websites if the content was not personalized, and soon users would lose their interest. Here comes the recommendation system for the rescue. In terms of volume recommendation system is able to handle a big amount of data. Although there has been a good amount of research on this topic, but still this remains a topic of interest given the significance.

While looking for the similar projects, I noticed that most of them have not tried to consider the **"Cold Start Problem"** i.e. showing recommendation for a new user or new product that our system has no data available for. In this project, I expect to deal with that part of the problem. Apart from this, my plan is to do a literature survey on "Cold Approach Problem" in the recommendation system.

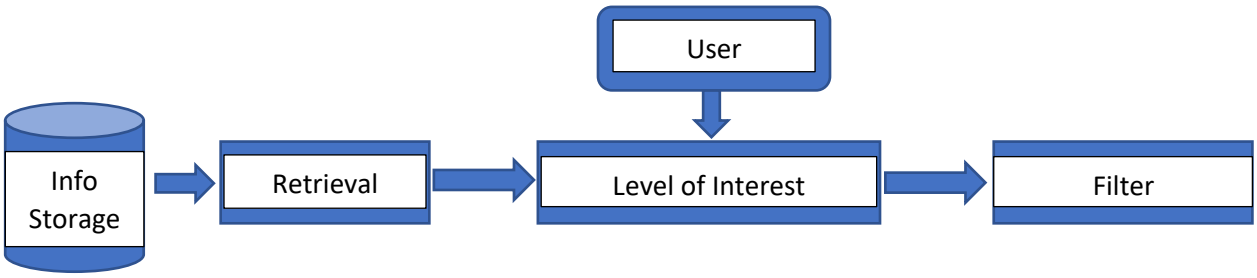
Recommendation System:

Recommendation system suggests items to a user based on his/her past behaviors, their personal data, their physical location or based on some form of similarity to others. I seem to date back the presence of commendation system in my life since I had my first mobile phone in 2004, it was capable to show me favorite contacts based on who I talk to more or message.

Therefore, a recommendation system is a system that is capable of collecting information, processing and showing personalized results to the user. Some most common examples of the recommendation system are: YouTube & Facebook news feed or friend suggestion, LinkedIn job suggestion or suggested TV series or movies on platform like HULU or Netflix.

As stated, that a recommendation system works on the data from the user e.g. their search pattern, visited places or pages etc. In many situations, system is unaware of

this information. That makes it very difficult for the system to predict the data for a new user. This is called cold approach problem.



Types of recommendation Systems:

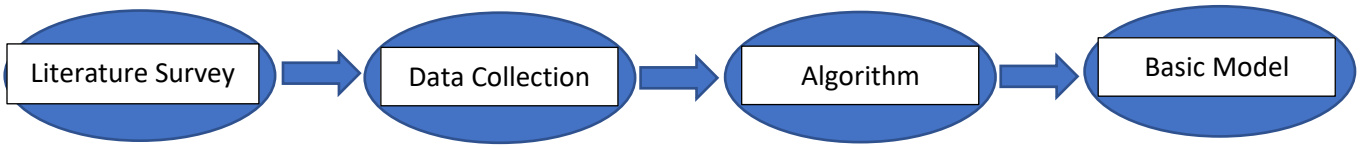
1. **Collaborative Filtering**
2. **Content Based Filtering**
3. **Demographic Filtering**
4. **Social Filtering**
5. **Hybrid Filtering**

Deliverables:

Most of the academic projects on this topic don't seem to hand the cold start problem. Hence, the purpose of this project is to first implement a basic recommendation based on collaborative filtering and address the Cold Start Approach. And at the end do a literature survey on Cold start approach and identify some new areas having scope for the research.

Milestone achieved: I have been steadily working on this project, so far, I have done following things.

1. Literature Survey on recommendation system
2. Dataset Collection
3. Algorithm Selection
4. Basic Model Creation for the recommendation system.



Dataset: For this problem I am using MovieLens 100K Dataset, this dataset contains 100,000 ratings from 1000 users on 1700 movies, released during 1998. The reason behind using this database is because in the database I suggested during the proposal, the geographic information was not given in the database, and I am going to use this information to deal with the cold start problem.

Link to the dataset : https://grouplens.org/datasets/movielens/100k/

Collaborative Filtering: Collaborative filtering is a technique that calculates similarity amongst users based on the ratings. It's a model that exploits user behavior and easy to implement, also it provides good results. There are 2 types of collaborative filtering: Memory Based collaborative filtering and Model Based collaborative filtering. Model based collaborative filtering is faster and better to use because it doesn't goes through the whole database every time.

Paper Selection for Literature Survey: Based on the comments received on the project proposal, I am planning to include a literature survey of cold start problem which is a big part of this project. I have identified the papers the initial papers[3][4][5][6] that I am going to use for the literature survey.

Algorithm Selection:

1. For this project, I am using kNN(k Nearest Neighbor) algorithm. It is one of simple and easy algorithms and gives good performance. This algorithm identifies the class this datapoint belongs to based on the distance from the other points.
2. **Cold Start Problem:** As mentioned earlier the reason behind picking up this database. The demographic information of the user is present in the dataset. I would be dealing using this information to deal with the new user cold start problem. [3][4][5][6] gives a good reference to deal with this problem, my approach in this project would be based on these papers.

a. The demographic information in this database looks something like this: **user id | age | gender | occupation | zipcode**
b. User ratings are given as follows: **user id | item id | rating | timestamp**

Model: Currently, I am working on to build the basic model of the recommendation system without addressing the cold start problem.

Milestone Pending:

1. Introduce the cold start problem in the implementation
2. Literature survey of cold start problem
3. Based on the literature survey, Identify some future problem
4. Observation the results and find a comparison

Checkpoints before Final Presentation: I have set up three checkpoints before the final presention.

1. **11/21/2018:** By this point, I want to finish the implementation part.
2. **11/28/2018:** I would be devoting this week to do the literature survey for the cold start problem.
3. **12/5/2018:** Then I plan on to spent rest of the time preparing report and presentation.

References:

1. https://machinelearningmastery.com/practical-machine-learning-problems/
2. https://www.ritchieng.com/machine-learning-project-customer-segments/
3. Callvik, Johan, Liu, Alva, (2017) "Using Demographic Information to Reduce the New User Problem in Recommender Systems." Kth Royal Institute of Technology School Of Computer Science And Communication

4. Manolis Vozalis, Konstantinos G Margaritis, (2004). “Collaborative Filtering Enhanced By Demographic Correlation.” AIAI symposium on professional practice in AI, of the 18th world computer congress.
5. MH Nadimi-Shahraki, M Bahadorpour, (2014). “Cold-start Problem in Collaborative Recommender Systems: Efficient Methods Based on Ask-to-rate Technique”. CIT. Journal of Computing and Information Technology 22 (2), 105-113
6. Rashid, A.M., Karypis, G., Riedl, J. (2002) “Learning preferences of new users in recommender systems: an information theoretic approach” 127–134. ACM Press
7. https://repository.ihu.edu.gr/xmlui/bitstream/handle/11544/29088/Product%20Recommendation%20System%20Final%20Version.pdf?sequence=1