

**1. Briefly describe how your solution works in English (not in pseudo code).**

Ans) Initially, I am looking for similar users. As given data set as less users compared to product, I am using user-based algorithm. To find the similarities between users **Correlation coefficients** is used. Here to find relationship between user I am using **Pearson's correlation** coefficient formula. A value of 0 indicates that there is no association between the two variables. A value greater than 0 indicates a positive association; that is, as the value of one variable increases, so does the value of the other variable. Here Threshold is set to 0.1, to avoid outliers, any relation with coefficient equal or more than 0.1 is considered as the good relation. After finding the relation between users Weighted sum is used to predict user rating on the Item.

**2. Briefly describe your strategy for the cold start issue (i.e., predicting an element in the matrix that relates to a new user who never gives any preference rating before or to a new item that no user ever gives rating to this item before).**

We should have default rating by analyzing and calculating average rating of other known users and initialize new user with the default rating and then keep learning about the new user or new item and keep updating according to his interest and matching him with the other users.

**3. Give two other examples of using a solution to the matrix completion problem in the real-world other than recommender systems and be specific in the physical meanings of the two dimensions as well as the entries of the matrix.**

**a. Seismic Geophysics:**

For Seismic data Matrix completion can be applied. Seismic data provide a “time picture” of subsurface structure. **2-D reflection seismic data** provide cross-sectional views in both the dip and strike directions. Data on the lines are a mixture of both in-plane and out-of-plane reflectors(2 dimensions). Seismic data is often irregularly sampled along spatial coordinates, which is mostly caused by dead or severely corrupted traces, surface obstacles, acquisition aperture and economic limit. Seismic data regularization spatially transforms irregularly acquired data to regularly sampled data in order to subsequently display, process and interpret it. Seismic data interpolation and reconstruction is one particular case of data regularization. Nowadays, the recovery of missing traces has become a main issue in seismic research, where input data are already measured or observed in a regular grid and one needs to reconstruct the value in missing traces (i.e., empty bins). Here **2 dimensions are time-data and depth.**

**b. Matrix Completion in Weather Prediction:**

Weather Prediction can be done using matrix completion. For that lets consider weather station and observations over a period of time be the 2 dimensions. Details like temperature wind speed and direction be the values in the matrix.

**4. Read at least three papers on recommender systems published within last three years other than the one you implemented in this project if you elected to take an existing solution from the literature, and write up a mini-survey for the papers you have read.**  
**Paper 1:**

**collaborative web recommendation systems based on and effective fuzzy association rule mining algorithm(farm).** A.KUMAR Research Scholar, Department of CSE, Sathyabama University, Chennai.

- Here author, proposed paper gives a method of developing a collaborative web recommendation system using Fuzzy Association Rule Mining. Fuzzy Association Rule Mining is the problem of discovering frequent item-sets using fuzzy sets in order to handle the quantitative attributes in transactional and relational databases.
- The proposed approach involves usage of fuzzy healthy association rule mining algorithm for Association Rule Mining. This approach helps the user to obtain the web sites which are most relevant to them.
- Recommendation using association rules is to predict preference for item k when the user preferred item i and j, by adding confidence of the association rules that have k in the result part and i or j in the condition part. This association rules can be generated using apriori algorithm. Association rules capture relationships among items based on patterns of co-occurrence across transactions.
- For fuzzy association rule mining standard ARM algorithms can be used or at least

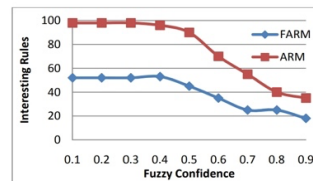


Figure 2: Number of Interesting Rules comparison

adopted after some modifications.

- The Fuzzy ARM Algorithm belongs to the breadth first traversal family of ARM algorithms, developed using tree data structures and it works in a fashion quite similar to the Apriori algorithm.

## Paper 2:

Self-tuning approach for implementing a multidimensional recommendation system using PID (Saurabh Chaudhari; Aditya Sawant; Reeta Patil) Publisher: IEEE.

- Here is a newly proposed algorithmic multidimensional approach that can be deployed to improve recommendation systems performance.
- PID is a closed loop, self-tuning algorithm which is predominantly implemented in mechatronics instruments where manual supervision is not feasible and it functions for correction of errors quantifiable in physical measures.
- The hybrid approach combines content-based and collaborative methods in order to mitigate the problems faced by the two filtering techniques individually.
- With its three-term functionality covering treatment to both transient and steady-state responses, Proportional-Integral-Derivative (PID) control offers the simplest and yet most efficient solution to many real-world control problems.
- **Proportional term (Kp)** - responsible for overall control action which is proportional to the error signal. **Integral term (Ki)** - responsible for reducing the steady state errors through compensation by an integrator. **Derivative term (Kd)** - responsible for improving response through compensation by a differentiator.
- The system exhibits multi-dimensionality by providing various parameters of items such as price, ratings, discount, etc. thereby generating a more comprehensive

recommendation list with optimum performance as it is coupled with PID

$$u(t) = \overbrace{K_p e(t)}^{\text{Proportional}} + \overbrace{K_i \int_0^t e(\tau) d\tau}^{\text{Integral}} + \overbrace{K_d \frac{d}{dt} e(t)}^{\text{Derivative}}$$

algorithm.

### PAPER3:

S. Amara and R. R. Subramanian, "Collaborating personalized recommender system and content-based recommender system using TextCorpus," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 105-109, doi: 10.1109/ICACCS48705.2020.9074360.

- This paper presents an effective approach for efficient retrieval of data of users' interest. Proposed the tagging system for the dataset and the new user-profile design in this paper.
  - To make data retrieval and acquisition efficient have proposed the inclusion of text corpus in the dataset. Deployment of the data from the dataset can benefit the auto-taggers. This model initially gets the crude data of the user. Depending on the activity of the user, the profile is updated and the data is recommended to the user.
  - Scrapy is an open-source python framework which is used to scrape the data, including image, video, text data, from the web page in a fast, simple and yet extensible way. This framework is used to scrape the visible content of the web pages such as articles, headings, URLs.
  - By using this the data in the websites such as Cnet and other technical or news websites data can be directly retrieved to the database. This data is further processed through the tag allocation process of NLTK.
  - NLTK is a platform that works with the help of human language leveraging features like sentence tokenizer and POS taggers. Python has an initial tokenizer but NLTK is more reliable and versatile. It will tokenize the sentences preprocessed with the removal of unnecessary prepositions and connectors {'this', 'is', 'a', 'token'}.
  - The process of Gibbs sampling employed in proposed framework is described.
  - Gibbs Sampling: Dynamic Model of the User Profile, Construction of UP-Tree, T-UP-Tree Algorithm is used in this paper
  - The tags of the user will be periodically updated and the relevant information is recommended to the user every time the user logs in.
5. Based on the literature you have read as well as your experience in this project, discuss what the open issues (other than cold start) are on the topic of recommender systems and how people are attempting to address these issues or these issues are purely open with no one ever attempted to address them.
- **Scalability:** Generally algorithms are more accurate with fewer data points. As the number of users grow, the algorithm suffers Scalability issues.
  - **Deficiency of information:** It takes a lot of cleared data to create a recommendation system that works efficiently and makes precise suggestions.

- **Information Variability:** information may be vague sometimes it can be changing.
- **Over Specialization:** A good recommender should suggest diverse items in which content-based system lacks. It hinders the user from discovering new and different.

By :

Name: SANJAY BETTADAPURA GANESHA

Bnumber: B00816568