# (E-AggNN) Execution of Effective and Efficient Algorithm using Collaborative Filtering and Rank Generation Similarity Search on Recommender System

Ms. Sonali Patil[1], Ms. Swati Joshi[2]

*M.E. in Computer Science & Engineering Pune University, Pune [1]*
*Pursuing M.E in Computer Science & Engineering Department BSIOTR, Pune University[2]*

**ABSTRACT**

*Recommendation systems are collaborative software that can be applied to expertise locating. A recommendation system that suggests people who have some expertise with a problem holds the promise to provide, in a very small way, a service similar to that provided by key people. Expertise recommendation systems can reduce the load on people in these roles and provide alternative recommendations when these people are unavailable. The architecture is open and flexible enough to address different organizational environments. Many real-world applications require solving a similarity search problem where one is interested in all pairs of objects whose similarity is above a specified threshold. We examine alternatives for incorporating feedback into the ranking process and explore the contributions of user feedback compared to other common web search features*.

**Keyword**: *Collaborativefiltering, HybridRecommendationsystem, Rank Generation, Aggregation*

## I. Introduction

In the field of Information Retrieval , document clustering is used to automatically group the document that belongs to the same topic in order to provide user's browsing of retrieval results [2]. Some experimental evidences show that IR application can benefit from the use of document clustering [3]. Document clustering has always been used as a tool to improve the performance of retrieval and navigating large data. With time getting passed, people face difficult problems that they cannot solve alone. In these situations the right people are the ones who have the expertise to answer a specific question or in some other way move the problem toward resolution. Using collaborative filtering to generate recommendations is computationally expensive [2]. It is $O(MN)$ in the worst case, where $M$ is the number of customers and $N$ is the number of product catalog items, since it examines $M$ customers and up to $N$ items for each customer. However, because the average customer vector is extremely sparse, the algorithm's performance tends to be closer to $O(M + N)$. Scanning every customer is approximately $O(M)$, not $O(MN)$, because almost all customer vectors contain a small number of items, regardless of the size of the catalog. But there are a few customers who have purchased or rated a significant percentage of the catalog, requiring $O(N)$ processing time. Thus, the final performance of the algorithm is approximately $O(M + N)$. Even so, for very large

data sets — such as 10 million or more customers and 1 million or more catalog items — the algorithm encounters severe performance and scaling issues. Once the algorithm generates the segments, it computes the user's similarity to vectors that summarize each segment, then chooses the segment with the strongest similarity and classifies the user accordingly. Some algorithms classify users into multiple segments and describe the strength of each relationship. The complex and expensive clustering computation is run offline. However, recommendation quality is low.1 Cluster models group numerous customers together in a segment, match a user to a segment, and then consider all customers in the segment similar customers for the purpose of making recommendations. Because the similar customers that the cluster models find are not the most similar customers, the recommendations they produce are less relevant.

This approach assumes that there are no interactions between the underlying features producing the original web search ranking and the implicit feedback features [8]. We now relax this assumption by integrating implicit feedback features directly into the ranking process. We compared two alternatives of incorporating implicit feedback into the search process, namely re-ranking with implicit feedback and incorporating implicit feedback features directly into the trained ranking function [8]. Our experiments showed significant improvement over methods that do not consider implicit feedback. The gains are particularly dramatic for the top $K$=1 result in the final ranking, with precision improvements as high as 31%, and the gains are substantial for all values of $K$. Our experiments showed that implicit user feedback can further improve web search performance, when incorporated directly with popular content- and link-based features. An ANN can be defined as a highly connected array of elementary processors called neurons. A few successful recommendation systems rely on implicit opinions, rather than explicit ratings. CF systems have another weakness for expertise recommendation. Many recommendation systems have very specific architectures that are tailored to recommend their specific type of artifact. As well, these architectures commonly implement a single clustering algorithm for all participants and all artifacts. The architecture that we will present is capable of supporting a range of collaborative recommendation models. An organizationally specific implementation will demonstrate the flexibility in the architecture. The problem of finding all pairs of similar vectors arises in many applications such as query refinement for search engines and collaborative filtering [2].

Today we can find many techniques in searching user relevant objects. Imagine a portal of hotel searching system. Typically such portals are based on a faceted browser, a form search system or a full text search system. We compared our ranking methods over a random sample of queries from the search engine query logs [8]. The queries were drawn from the logs uniformly at random by token without replacement, resulting in a query sample representative of the overall query distribution. On average, 30 results were explicitly labeled by human judges using a six point scale ranging from "Perfect" down to "Bad". Overall, there were over 83,000 results with explicit relevance judgments. In order to compute various statistics, documents with label "Good" or better will be considered "relevant", and with lower labels to be "non-relevant".

## II. RELATED WORK

Older customers can have a glut of information, based on thousands of purchases and ratings. Customer data is volatile: Each interaction provides valuable customer data, and the algorithm must respond immediately to new information. A traditional collaborative filtering algorithm represents a customer as an *N*-dimensional vector of items, where *N* is the number of distinct catalog items [2]. The components of the vector are positive for purchased or positively rated items and negative for negatively rated items. The selling items, the algorithm typically multiplies the vector components by the inverse frequency (the inverse of the number of customers who have purchased or rated the item), making less well-known items much more relevant.

## III.LITERATURE SURVEY

Document Clustering using Concept Space and Cosine Similarity Measurement [1]: Document clustering is related to data clustering concept which is one of data mining tasks and unsupervised classification. It is often applied to the huge data in order to make a partition based on their similarity. Initially, it used for Information Retrieval in order to improve the precision and recall from query.

Recommendations Item-to-Item collaborative Filtering [2]: Recommendation algorithms are best known for their use on e-commerce Web sites,1 where they use input about a customer's interests to generate a list of recommended items.

Search system on collaborative filtering [3]: The meaning of goodness can be for each user (or group of users) different. The natural meaning of this ordering is that the user determines the relations "'better" or "worst" between any two values of a given property.

An artificial neural network [4] : Electric Load Forecasting Using An Artificial Neural Network D.C. Park, M.A. El-Sharkawi, R.J. Marks II, L.E. Atlas & M.J. Damborg.

A model of trust based recommendation system on social network[5]: Frank E.Walter,Stefano battiston and frank Schweitzer.

A Flexible Recommendation System and Architecture [6]: David W. McDonald and Mark S. Ackerman Expertise Recommender,December 2-6, 2000, Philadelphia, PA.Copyright 2000 ACM 1-58113-222-0/00/0012…$5.00

Scaling Up All Pairs Similarity Search [7] : Roberto J. Bayardo Copyright is held by the International World Wide Web Conference Committee (IW3C2).

Improving Web Search Ranking by Incorporating User Behavior Information [8] : Eugene Agichtein SIGIR'06, August 6–11, 2006, Seattle, Washington, USA.Copyright 2006 ACM 1-59593-369-7/06/0008.

## IV.PROBLEM DESCRIPTION

Collaborative filtering (CF)-based recommender systems represent a promising solution for the rapidly growing web market. However, in the Web environment, a traditional CF system that uses explicit ratings to collect user preferences has a limitation: customers find it difficult to rate their tastes directly because of poor interfaces and

high telecommunication costs. Implicit ratings are more desirable for the Web, but commonly used cardinal (interval, ratio) scales for representing preferences are also unsatisfactory because they may increase estimation errors. In this paper, we propose ae-AggNN -based recommendation methodology based on both implicit ratings and less ambitious ordinal scales.

## V.PROPOSED SYSTEM

The main contributions of this paper are:

- A broad exploratory assessment, appearing that our calculations can create query results with great precision Two approximation algorithms for the ANN issue, one for the addition variant and the other for the max variant.
- A hypothetical investigation of our strategies, indicating conditions under which a careful result can be ensured. We likewise indicate conditions under which rough result can be ensured for a variable separation estimate proportion.

Therefore, using the item ratings and user profiles, recommender system has been proposed to provide diverse recommendations i.e. highly personalized items with only a minimal accuracy loss as well as suggest a sequence of items instead of a single recommendation to improve the quality of recommendations and use consumer-oriented or manufacturer oriented ranking mechanisms so both consumer and manufacturer will get benefit from recommendations [8].

## IV.SYSTEM DESIGN

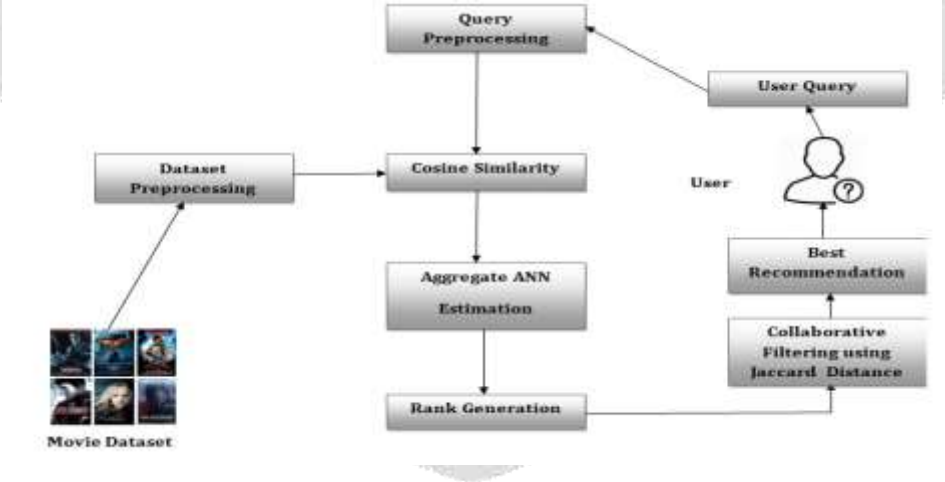


Fig 1: System Architecture

The proposed recommendation system the user may give query for the expected result [4][6]. It is given to then best recommendation system. After than in collaborative filtering using jaccard distance the given item is compared with the database available using the rank generation. Cosine similarity measurement as replacement of Euclidean distance to involve in fuzzy c-means [1]. Cosine similarity embedded to experiment evaluation. Implementing Fuzzy C-means clustering algorithm by term-document V vector as representative of the document collection and using Cosine similarity as replacement of Euclidean distance and can be applied into the objective function of Fuzzy C-means algorithm[1].

Traditional collaborative filtering, cluster models and search-based methods. Here, we compare these methods with our algorithm, which we call item-to-item collaborative filtering. Unlike traditional collaborative filtering, our algorithm's online computation scales independently of the number of customers and number of items in the product catalog. Our algorithm produces recommendations in realtime, scales to massive data sets, and generates high quality recommendations. There are three common approaches to solving the recommendation problem: traditional collaborative filtering, cluster models, and search-based methods. Here, we compare these methods with our algorithm, which we call item-to-item collaborative filtering. Unlike traditional collaborative filtering, our algorithm's online computation scales independently of the number of customers and number of items in the product catalog. Our algorithm produces recommendations in real-time, scales to massive data sets, and generates high quality recommendations.

Two popular versions of these algorithms are collaborative filtering and cluster models. Other algorithms including search-based methods and our own item-to-item collaborative filtering focus on finding similar items, not similar customers. For each of the user's purchased and rated items, the algorithm attempts to find similar items. It then aggregates the similar items and recommends them.

Using collaborative filtering to generate recommendations is computationally expensive. It is O(MN) in the worst case, where M is the number of customers and N is the number of product catalog items, since it examines M customers and up to N items for each customer. However, because the average customer vector is extremely sparse, the algorithm's performance tends to be closer to O(M + N). Scanning every customer is approximately O(M), not O(MN), because almost all customer vectors contain a small number of items, regardless of the size of the catalog.

Item-to-item collaborative filtering, scales to massive data sets and produces high-quality recommendations in real time. Item-to-item collaborative filtering matches each of the user's purchased and rated items to similar items, then combines those similar items into a recommendation list.

The key to item-to-item collaborative filtering scalability and performance is that it creates the expensive similar-items table offline. The ANN is used to learn the relationship among past, current and future views for the given choices. Recommendation systems are collaborative software that can be applied to expertise locating. This work describes a general recommendation architecture that is grounded in a field study of expertise locating. Recommendation systems are one possible technology that can augment and assist the natural expertise locating behavior in organizations. A recommendation system that suggests people who have some expertise with a problem holds the promise to provide, in a very small way, a service similar to that provided by key people. Systems that assist with expertise location are similar to a broad class of systems known as recommendation systems [4].

Recommendation systems are commonly used to recommend documents based on user preferences. This definition distinguishes recommendation systems from systems that are more properly characterized as information retrieval or information filtering. Recommendation systems are not completely synonymous with collaborative filtering (CF) systems. Many CF systems rely on explicit statements of user opinion, such as ratings, to create user profiles. By

relying on ratings, CF systems often have difficulty generating the initial user profile and, as the profiles develop, must deal with a sparseness of ratings relative to the total number of items. These are two active areas of CF research. A few successful recommendation systems rely on implicit opinions, rather than explicit ratings

In many recommendation systems, profiles are a list of items which an individual has rated. These ratings profiles are used in two ways. First, profiles are clustered to create groups of users who have similar likes and dislikes. and secondly profiles are used to identify items that a user has not yet rated and are therefore good candidates for recommendation.

**VI. RESULTS AND DISCUSSIONS**

The proposed methodology of recommendation using aggregate nearest neighbor methodology is using movie dataset taken from IMDB. Which is extracted from following URL https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset .

This section is evaluating the performance of the cosine similarity that is implemented to extract the values of aggregate nearest neighbors. System is deployed using java based windows machine using Netbeans as standard development IDE.

To show the effectiveness of the system some experiments are conducted using Mean absolute error (MAE) . MAE is used to finding the error that eventually occurred during the process of cosine similarity while computing Aggregate nearest neighbor for recommendation.

MAE can be representing from the equation 1, which is broadly using to measure the prediction quality of recommendation systems.

$$MAE = \sum_{i,j} | r_{i,j} - r'_{i,j} | / N \quad \text{---------(1)}$$

Where, *ri,j* denotes the expected similarity from the cosine similarity rules for $i^{th}$ query from the user with $j^{th}$ similarity measured value.

*r'i,j* denotes the calculated similarity from the cosine similarity rules for $i^{th}$ query from the user with $j^{th}$ similarity measured value.

We use the Normalized Mean Absolute Error (NMAE) metric to measure the prediction quality of our cosine similarity method that is deployed in hybrid Recommendation process. We define our NMAE to be the standard MAE normalized by the mean of the expected similarity measure values as follows:

$$NMAE = MAE / (\sum_{i,j} r_{i,j} / N ) \quad \text{--------(2)}$$

Where smaller NMAE value means higher prediction quality. different trials of our experiment it yields different NMAE as listed in table 1.

| No. Runs | Hybrid Recommendation | Collaborative Filtering Recommendation |
|---|---|---|
| 1 | 0 | 0.9 |
| 2 | 3.6 | 4.5 |
| 3 | 2 | 8 |

| | | |
|---|---|---|
| 4 | 1.8 | 2 |
| 5 | 1.88 | 8 |
| 6 | 2.8 | 16.3 |

**Table 1: Comparative NMAE Values**

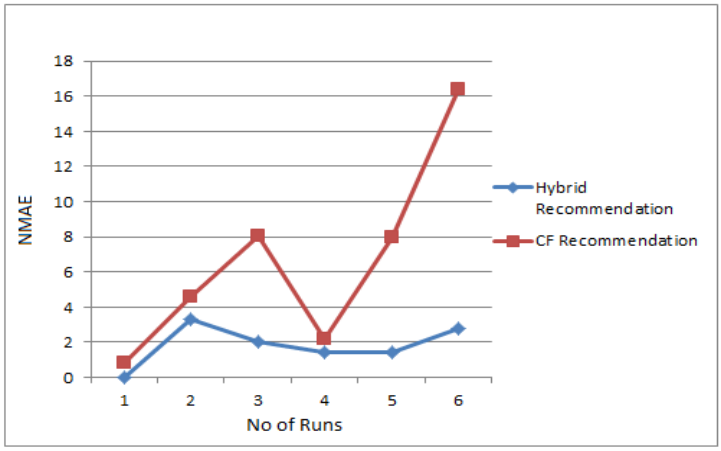


**Figure 1: Performance Graph**

The plot in the figure 1 clearly indicated that hybrid recommendation of our system clearly indicates the lesser NMAE, This means our system over performs than the traditional collaborative filtering.

## VII. CONCLUSIONS

Recommender systems provide valuable suggestions to users with the help of user rating databases. Therefore user rating database creation is one important step in the proposed system. Database creation module is implemented by accepting explicit ratings from the different users. Similarity computation module computes similarity between target item and other items while prediction estimation module predicts the rating for the target item. Accuracy of predictions can be measured with statistical accuracy metricsFurther, the system proposed recommendation technique which is based on content i.e. user preferences and item profiles. Item popularity based parameterized ranking technique will ranks the items such that recommendation accuracy will be maintained and the diversity will be increased. Quality of recommendations will be improved using consumer/ manufacturer oriented ranking and item sequence generation techniques.

## VIII. ACKNOWLEDGEMENT

I take this special opportunity to express my sincere gratitude towards professor and all the people who supported me during my entire project work. I would like to express my gratitude to my guide and also the project coordinator Prof. Sonali A.Patil for providing special guidance. I would also like to thank our HOD Prof. G. M. Bhandari who

always has enough time to solve student's problems at any time. Finally thanks to all teachers who are always supportive at us.

## IX. REFERENCES

[1] S. Lailil Muflikhah, Baharum Baharudin, Document Clustering using Concept Space and Cosine Similarity Measurement. 2009 International Conference.

[2] Greg Linden, Brent Smith, and Jeremy York, Item-to-Item Collaborative Filtering. IEEE Computer Society 1089-7801, JANUARY • FEBRUARY 2003

[3] User preference based search system Peter Gurský[1], Tomáš Horváth[1], Róbert Novotný[1], Veronika Vaneková[1], University of P.J. Šafárik, Košice, Slovakia

[4] Electric Load Forecasting Using An Artificial Neural Network D.C. Park, M.A. El-Sharkawi, R.J. Marks II, L.E. Atlas & M.J. Damborg,"Electric load forecasting using an artificial neural network", IEEE Transactions on Power Engineering, vol.6, pp.442-449 (1991).

[5] A model of trust based recommendation system on social network, Frank E.Walter,Stefano battiston and frank Schweitzer.

[6] David W. McDonald and Mark S. Ackerman Expertise Recommender: A Flexible Recommendation System and Architecture December 2-6, 2000, Philadelphia, PA.Copyright 2000 ACM 1-58113-222-0/00/0012…$5.00

[7] Scaling Up All Pairs Similarity Search Roberto J. Bayardo Copyright is held by the International World Wide Web Conference Committee (IW3C2).

[8] Improving Web Search Ranking by Incorporating User Behavior Information Eugene Agichtein SIGIR'06, August 6–11, 2006, Seattle, Washington, USA.Copyright 2006 ACM 1-59593-369-7/06/0008.

**X.BIOGRAPHIES**

| | |
|---|---|
|  | Ms.Swati Shripad Joshi.<br>Pursuing M.E in Computer Science & Engineering<br>Department BSIOTR, Pune University |
|  | Prof. Sonali Appasaheb Patil<br>Mtech CSE, Phd pursuing from BSAU, Chennai.<br>Asst.prof.Department of Computer Engineering<br>JSPM'S BSIOTR,WAGHOLI,PUNE |