# Fusion of Similarity Measures in Recommender Systems

**Ashishkumar Patel[1], H B Bhadka[2]**

[1]Assistant Professor, LDRP ITR, KSV University, Gandhinagar, Gujarat, India
PhD Scholar, C U Shah University,Wadhwan, Gujarat, India
[2] Dean, Faculty of Computer Science, C U Shah University, Wadhwan, Gujarat, India
Email: abp5@live.com

Received: 16.04.2024 Revised : 15.05.2024 Accepted: 25.05.2024

**ABSTRACT**
Recommender systems provide suggestions for goods (such as movies or products) that a user may likely enjoy, taking into account their past preferences. An essential component of these systems is computing the similarity scores among users in order to detect individuals who share similar preferences. In order to enhance the precision of computing user similarities, this work introduces a novel "Fusion of Similarity Measures in Recommender Systems". Contemporary similarity metrics such as Pearson correlation exhibit certain limitations. Sometimes, they exhibit a strong correlation between users despite significantly varied ratings, or a weak correlation for users with almost identical ratings. Furthermore, they fail to adequately consider variables such as the proportion of items that are co-rated by two users. Our approach computes user similarity in a novel manner that specifically tackles these concerns. It incorporates many parameters including rating closeness (how near the ratings are), rating significance (how distant from the average rating), rating singularity (the uniqueness of the ratings), and proportion of co-ratings. It also adapts for individual users' rating tendencies. Experiments demonstrate that our approach yields more precise similarity ratings then current approaches. It accurately classifies users as having high or low similarity depending on the measurable level of preference alignment. This significantly improves the quality of suggestions provided to users by the recommender system.

**Keywords:** Recommender System, Item Similarity, User Similarity, Similarity Fusion

## 1. INTRODUCTION

Many e-commerce websites provided a large selection of goods to their customers in the late 1990s dot-com boom. These goods can number in the millions and their quantity varies from site to site. Each company was more likely to have a product that was appropriate for a certain user since there were more things available, but users found it challenging to distinguish between acceptable products due to the sheer volume of options. greater challenge (Ricci et al., 2015) [1].

E-commerce businesses started suggesting items from a small catalog in order to solve this issue (Ekstrand et al. 2011b) [2]. We also offer a list of recommendations based on past purchases and what the user was looking for when the recommendation was generated, but consumers still need to browse the complete inventory of things. These e-commerce sites generated recommendations by utilizing technology including databases, graphical user interfaces, and algorithms. These systems were first referred to as collaborative filtering by researchers (Goldberg et al., 1992) [3]. Nevertheless, scholars subsequently embraced the phrase "recommender system" extensively (Resnick and Varian, 1997) [4].

Because recommendation algorithms are so common on e-commerce sites, they have drawn a lot of attention (Ricci et al., 2015) [1]. However, the purpose of these systems was to assist users in navigating the information space rather than to help them make purchases from online stores or boost sales (Rich, 1979; Goldberg et al., 1992) [3]. The recommendation algorithm suggests a range of content that may be seen online and doesn't require further purchase, including songs, movies, articles, and more. Items are defined by Lissi et al. (2015) as any real or virtual objects, goods, services, or procedures that are recommended to you by online, email, or text message recommendation systems. Products from Flipkart, YouTube videos, and Spotify music are a few examples of things.

## 2. Review Findings of Literature Survey

Systems that make recommendations are essential for pointing users in the direction of pertinent information, goods, or services. Pure recommender systems do have certain drawbacks, though. Researchers have looked into hybrid techniques, which mix various strategies, to overcome these issues. These hybrid

approaches present a viable way to improve the quality of recommendations while addressing intrinsic limitations. Here, four distinct approaches become apparent:
Algorithm Separation and Result Aggregation: We may capitalize on the advantages of each strategy by executing different recommendation algorithms separately and combining the outcomes.
Guidelines for Integrating Content-Based Filters: Respecting certain rules when merging collaborative and content-based filtering methods guarantees a smooth integration. Collaborative filtering depends on human involvement, whereas content-based filtering makes use of item attributes.
Content-Based Strategies with Collaborative Filtering: We may use user-item interactions while taking item features into account by using collaborative filtering techniques within a content-based framework.
Unified Recommender Systems: Building a single system that combines collaborative and content-based filtering techniques in a smooth manner offers a complete answer.
Several key conditions apply to the successful deployment of recommender systems across various algorithms:
Rich Item Space: Domains frequently have a large number of items, which prevents users from doing a thorough exploration. Recommender systems fill this need by perceptively recommending pertinent options depending on user preferences.
Preference-Driven Decisions: The choices made by users are heavily influenced by their preferences. The usefulness of a recommender system is its capacity to match subjective preferences, particularly in situations where objective criteria are insufficient.
Taste Data Repository: The foundation of successful recommendations is taste data, which is gathered from both explicit and implicit user actions and evaluations. Taste data describes how consumers engage with products.
Items in a domain have common attributes, which is known as homogeneity of items. The system can reliably use taste data for viewing, rating, and other interactions because to these shared properties.
In conclusion, adopting universal principles and hybrid methodologies enable recommender systems to provide context-aware, individualized recommendations that improve user experiences in a variety of contexts. We can get around some of the drawbacks and issues with pure recommender systems by using hybrid techniques. the combination of methods may work in various ways [6]:
1) Implementing algorithms separately and combining the outcomes.
2) While collaborating, follow some content-based filtering principles.
3) When employing a content-based approach, follow basic rules for collaborative filtering.
4) Build a single recommender system that integrates the two approaches.
A few requirements must be met for a recommender system to be implemented successfully. Regardless of the recommender algorithm, there are some applicability that hold true for every recommender [5].
Numerous items There are numerous aspects of domains that people could find interesting. The user is unable to go through each one of them. Decisions based on personal tastes: What each user chooses depends on their preferences. In the event that you have objective standards for suggesting items to readers, a recommender system is not particularly helpful.
flavor Data: The system should record user interactions with items that may be taken to be flavor signals. Whether this is an implicit compilation of user activities, which ought to be included, or explicit rating data is inconsequential.
Items that are homogenous: The domain's items share some characteristics, such as the ability to all be viewed or evaluated and the ability to be covered by taste data.

## 2. Motivation and Objectives

In general, recommendation algorithms look for higher accuracy by using item kinds such popular things (Smith and Linden, 2017; Lu et al., 2012) [7] and items that match those in the user's profile (Iaquinta et al., 2010). In this work, popularity is determined by topic ratings on a certain numerical value or recommendation system. These objects in the system are usually popular items found in portals or applications, like books for sale (Salma Gerada, 2009). Item resemblance can be determined by looking up user ratings or by using the item's characteristics. If two things have similar characteristics, they are deemed comparable. B. When two films are highly rated by the same users or fall into the same genre. Popular products are frequently given better ratings because of their superior quality (Kotkov et al., 2017; Selma Guerrada, 2009).
The accuracy of a recommender system is usually a crucial statistic, but McNee et al. (2006) state that this does not guarantee that a user or client will be satisfied. He gets a kick out of reviewing products and finding out how much others liked them, but he doesn't know if they liked the recommendations. Regardless of recommendations, users can use and appreciate the goods they are familiar with, and they can receive precise recommendations. These recommendations, nevertheless, frequently fall short of what users re-

quire. They must locate items they are interested in but are unsure of or unable to locate (Herlocker et al., 2004; McNee et al., 2006).
The goal of this study is to find and combine various similarities in order to raise the recommendation score for each item. The following are the research questions and the reasons behind them:
• What is a system of recommendations?
• Collaborative filtering based on both users and items.
• How to assess and assess RS parameters.
• Variations in similarity metrics.
• Models of similarity are compared.
• In RS, what is serendipity?
• Elements of chance encounters.
• How serendipity is quantified.
• Action items for coincidental RS.
• Definition of Similarity Fusion.
• Evaluations against alternative similarity models.

## 3. Dataset

The University of Minnesota's GroupLens Research Project provided the MovieLense [8] data sets. 1682 movies had been rated by 943 users in this sample. Each movie's rating, which ranges from 1 to 5. There are 100,000 ratings in total in this dataset. Information about each user and their movie rating of at least 20 films is contained in this dataset. The University of Minnesota's Computer Science Department is in charge of gathering it. The data in the dataset is tab separated and arranged arbitrarily.
item_id, timestamp, rating, and user_id.
We used the dataset above to test a number of algorithms. The output of this implementation will list the top n items together with their item ids.

## 4. Experimental Setup

Step 1: The application reads data into a 2D array from a file called theMovieLense Dataset, which contains user-item ratings.
Step 2: Uses a variety of techniques, including Pearson, JGR, PEARJAG, URP, and Fusion (Our Algorithm), to calculate user similarity values.
Step 3: Utilizing the estimated similarity values, determines the user's closest neighbors.
Step 4: Determines the weighted average of reviews from the user's closest neighbors in order to recommend products.

## 5. RESULTS AND DISCUSSIONS

By computing similarity utilizing a similarity measure, our Fusion technique is utilized in the suggested architecture to get over the limitations of the PCC measure (cold start condition and normalization).
The top value is determined by a group of common users, and the result is adjusted top-k. We discover the top k neighbor (user) using the k-nn approach (use the modified top k in-stead of top k).
The shortcomings of the current similarity metrics are demonstrated here. A novel similarity measure method based on fusion approach is provided to address these shortcomings. Moreover, the percentage of ratings that two users share is taken into account by the improved similarity metric. Because different users have different rating preferences, we used the rating's mean and variance to describe each user's rating preference (URP). These outcomes demonstrate the effectiveness of our method and its capacity to get around the drawbacks of the current similarity measurements.

## 6. CONCLUSION AND FUTURE WORKS

### Conclusion

The majority of techniques encounter issues due to their high levels of complexity and ambiguity. Current studies in the field go beyond the system's accuracy. Personalized services for users are increasingly being provided through collaborative filtering. When there aren't enough ratings to determine the similarities between each user, we'll employ a new user similarity model to enhance suggestion performance. The model takes into account both the global preference of user behavior and the local context information of user ratings. We examined the shortcomings of the current measures of similarity. Because different users have different preferences for ratings, we will use the rating's mean and variance to explain each user's preference. These outcomes demonstrate the efficacy of our fusion approach and its capacity to get around the drawbacks of the current similarity measurements.

**Future Scope**

Even though some of these issues will likely be resolved in the future, our fusion strategy is likely to remain a crucial component of recommender systems. Combining our method with other approaches like matrix factorization or deep learning to create a hybrid recommender system that can handle more intricate data patterns and produce recommendations that are better is a potential strategy.

Future studies could also look into novel similarities in an effort to better understand the preferences and behaviors of target users. Contextual data, such as time, place, or social network, can enhance the precision and applicability of recommendations.

Ultimately, the need to create algorithms that can provide tailored suggestions without depending on private user data is growing as privacy concerns do. Models may be trained on decentralized data without compromising privacy thanks to techniques like differential privacy or amalgamated learning. With continuous research targeted at improving its accuracy, scalability, and privacy, the future is bright overall.

**REFERENCES**

[1] Ricci, F., Rokach, L., & Shapira, B. (2015). Introduction to recommender systems handbook. Springer.
[2] Ekstrand, M. D., Riedl, J. T., & Konstan, J. A. (2011). Collaborative filtering recommender systems. Foundations and Trends® in Human–Computer Interaction, 4(2), 81-173.
[3] Goldberg, D., Nichols, D., Oki, B. M., & Terry, D. (1992). Using collaborative filtering to weave an information tapestry. Communications of the ACM, 35(12), 61-70.
[4] Resnick, P., & Varian, H. R. (1997). Recommender systems. Communications of the ACM, 40(3), 56-58.
[5] M. Thesis, “Petr Cvengroˇ s Universal Recommender System,” 2011.
[6] X. Su and T. M. Khoshgoftaar, “A Survey of Collaborative Filtering Techniques,” Adv. Artif. Intell. vol. 2009, no. Section 3, pp. 1–20, 2009.
[7] Brent Smith, Greg Linden, Two Decades of Recommender Systems at Amazon.com, IEEE Internet Computing, 2017
[8] MovieLense, http://www.movielens.org/