HYBRID APPROACH FOR MOVIE RECOMMENDATION SYSTEM USING DOUBLE COLLABORATIVE FILTERING AND SEQUENTIAL PATTERN ANALYSIS

HARRY YANTO [1], SANI MUHAMAD ISA[2]

[1,2]Computer Science Department, Binus Graduate Program – Master of Computer Science Bina Nusantara University, Indonesia

E-mail : [1]harry.yanto@binus.ac.id, [2]sani.m.isa@binus.ac.id

ABSTRACT

Recommendation systems (RecSys) are essential nowadays to handle information overloading. RecSys algorithms usually predict whether the user will like the content or not based on the previous consumed content by the user. User's preferences will be changing from time to time, but the research of RecSys algorithm nowadays rarely involves the sequence of user preferences. We also trying to engage two Collaborative Filtering algorithms (CF) with different attributes. CF that use the rating as attribute (CFR), and CF that use the user-preferred genres as attribute (CFG). We hybrid those three algorithms CFR, CFG, and Sequential Pattern Analysis (SPA) to increase the accuracy of the recommendation system. Then we evaluate it using the f1 score and compared it to the CFR, CFG, and SPA alone. This research concludes that hybrid CFR, CFG, and SPA increase the accuracy of f1 and precision score compared to the algorithm stands alone. We also conclude that cosine is the best similarity to use in searching similar users for RecSys

Keywords: *Recommendation System*, *Hybrid, Collaborative Filtering*, *Sequence Pattern Analysis*, *Best Similarity*

1. INTRODUCTION

In the era of digitization, Movies on Demand services simplify the consumption of digital content such as movies. Owning a vast quantity of digital content enables the company to generate a substantial profit. On the other side, because each user has unique preferences, the business must deliver relevant material. This produces information overload problem [1–3]. The user may lose interest in browsing the website's content as a result of an overload of information. Therefore, a recommendation system (RecSys) is required.

RecSys originally purposed by GroupLens for Usnet News [4], was a system that analyses customer's purchase history to identify their purchase patterns [5] to provide personalized recommendations [6]. RecSys are not lika a search engine that provides result based on semantic matching results, but it helps to filter information shown to the user based on the user's preferences, so the information stays relevant for each unique user[7–9]. RecSys also brings benefits to the service provider [10], by providing the right information to the user, it encourages people to spend more time on the website [11] and eventually purchase from your website [1] which leads increased sales growth and establish customer loyalty [12, 13]. Therefore, all businesses are attempting to establish their own RecSys to enhance their targeted marketing campaigns. [14, 15].

Huang et al, [1] succeed to improve the quality of their recommendation by combining SPA and CF algorithm. By combining multiple algorithms with multiple attributes, we found out that the recommendation result become more accurate. In this paper, we are trying to experiment whether adding more algorithm with another attribute will improve the quality of the recommendation. We are trying to extend their algorithm with another CF algorithm that uses genres as an attribute. So, we are going to hybridize three algorithms, which is two CF algorithms with distinct attributes: attribute of genres (CFG) and CF with the attribute of ratings (CFR), along with SPA algorithms, using weighing technique to evaluate whether this will yield more accurate recommendations.

Choi et al [16] experimenting with 3 distinct correlation equation in algorithm CF to determine the best equation that produced recommendation.

Meanwhile there is another equation that's supposedly works best in helping CF algorithm to finds similarity between users, which is Jaccard similarity equation. Building on that research we are conducting an experiment to compare Jaccard similarity with others similarity equation used in those research to determine which of four different similarity equations (Cosine similarity, Euclidean distance, Jaccard similarity, and Pearson correlation) should be implemented into CFG and CFR, as well as how much weight should be assigned to each algorithm to produce the optimal recommendation.

2. RELATED WORKS

Based on survey that conduct by Sunilkumar, et al [17], there are many ways to predict user's ratings to achieve the best recommendation. Collaborative Filtering (CF), Content-Based Filtering (CBF), and Hybrid Method Filtering are the top three most used techniques to produce rating prediction. CF and CBF works by analyzing user's given rating and use similarity equation to find the similar item or user preferences in order to predict the next item or users that should be recommended. Hybrid Method on the other hand are method of combining various technique to achieve better rating prediction or recommendation list.

2.1. Collaborative Filtering

Collaborative Filtering (CF) is one of the most popular algorithms in RecSys; it is a method that generates recommendations based on a user's historical behavior and preferences, as well as the collaborative aspect of each user [4]. Some of the famous CF implementations in RecSys include GroupLens [18], Tapestry [19], MovieLens [20], Ringo [21], and Siteseer [22]. There are many research recommendation systems that use CF as their based algorithm [23, 24]. One of the benefits of using the CF algorithm is that it can identify cross-genre niches and generates more relevant recommendations than other algorithms [25, 26], It is adaptive, meaning that it improves over time, and in some circumstances, it can be applied to implicit rating data. However, this algorithm suffers from cold start and spatial data issues [15] that are common these days. This algorithm would not be able to perform well if there were a small number of users who provided ratings or preferences regarding the content. Additionally, this algorithm would not recommend novel items or content which hasn't been consumed. Many are attempting to improve the accuracy of CF algorithms by combining them with other algorithms such as Content-Based Filtering (CBF) [27], Sequential Pattern Analysis (SPA) [16, 28], and others. However, studies that consider sequential consumption pattern as an attribute factor to predict rating are still poorly researched [29]. In this paper, we will also conduct an experiment to determine whether combining another CF with other attributes improves the accuracy of a recommendation system.

2.2. Sequential Pattern Analysis

SPA was first introduced by Agrawal and Srinkat [30] as a method for analyzing the user's consumption sequence history and using the number of supports as a filtering method to predict which items are likely to be consumed next. There are papers [1, 16, 28], that includes sequential as an attribute when making predictions. According to these studies, incorporating the SPA algorithm will increase the recommendation system's precision.

2.3. Hybrid Method Technique

In their paper, Huang et al. investigate the combination of SPA and CF, with SPA as one of the most important factors in determining the recommended items [1]. They are successful in developing a hybrid CF and SPA algorithm using the feature Augmentation technique. This method replaces the conventional CF algorithm, which has difficulty dealing with situations in which a customer's preferences changed gradually. First, they cluster the customer population based on the targeted customer using the genetic algorithm, then they mine the sequential pattern for each cluster, and finally, they predict the top-N items using the two-stage recommendation. On the basis of the f1 score, they concluded that the proposed method outperforms conventional methods when it comes to generating recommendations. However they didn't consider using user's genre preferences, which is important, as their attribute to produce recommendation.

Choi et al. proposed the Hybrid Online-Product rEcommendation (HOPE) system, which is a hybrid algorithm combining implicit collaborative filtering and sequential pattern analysis [16]. This hybrid algorithm demonstrated that the combination of CF and SPA produces a more accurate recommendation than CF or SPA alone. They are combining the item scores produced by each algorithm using weighing. In this experiment, different Similarities-equations such as Cosine, Pearson, and distance-based similarity are being evaluated. They conclude that cosine similarity is the best similarity measurement, so they use the cosine similarity equation to calculate the score for each item in the CF algorithm. This paper suggests that the combination of SPA and CF produces a more precise recommendation system. However they also didn't consider user's genre preferences to produce recommendation, also they didn't experimented on Jaccard similarity which is supposedly produce more distinct similarity score between users.

Using a weighing technique, Liu et al. [28] are also attempting to hybridize CF and SPA algorithms. Using precision, recall, and f1 score, they also assess the accuracy of their new algorithms. This study concluded that combining CF and SPA could increase the f1 score and generate more accurate recommendations. However this paper also didn't realise the important of considering genre attribute as a factor to determined similarity of the user.

Hybrid algorithms are already making a great improvement over the current state-of-the-art method. [1, 16, 28] However, those algorithms miss to include genre preferences as a factor to produce the recommendation list, hence this research conducted.

3. PROPOSED METHODOLOGY

This research will utilize the 1M-ML MovieLens Dataset [31] It includes 1,002,009 anonymous assessments of about 3,900 films made by 6,040 MovieLens members who joined in 2000.

This dataset is used by the majority of researchers experimenting with recommendation system algorithms, and it is the most adaptable and straightforward dataset accessible for researching recommendation systems.

Previous research has shown that the hybrid CF and SPA algorithm produces superior suggestions. We will increase the algorithm's depth in the hopes of achieving a better recommendation result by fusing another CF algorithm with film genre as the attribute (CFG) to an existing hybrid CF algorithm with ratings as the attribute (CFR) and SPA algorithms, employing a weighting approach, and evaluating the output with Precision, Recall, and F1 Score. Moreover, we conduct experiments on a variety of similarity equations, including Euclidean, Jaccard, Pearson, and Cosine to search for the best similarity to produce the best recommendation.

We will conduct this research by first doing pre-processing data that explained in chapter 3.1. After that, we will produce recommendation using every algorithm to see the performance of every algorithm by them self, for algorithm CFG amd CFR, we conduct experiment with the similarity equation candidates to see which similarity equation trumps

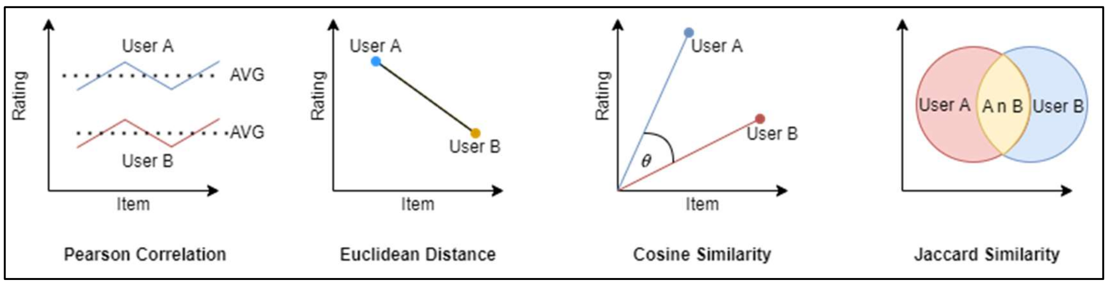

Figure 1: Perspective of each Similarity Equations

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1: Algorithm Results Table.

K-user	K-list	Sim	CFR_w	CFG_w	SPA_w	Algorithm	Eval	Score
13	80	Cos	-	-	-	CFG	F1	0,0662
13	80	Cos	-	-	-	CFG	Precision	0,1233
12	10	Cos	-	-	-	CFG	Recall	0,0667
11	80	Jac	-	-	-	CFR	F1	0,0736
12	80	Cos	-	-	-	CFR	Precision	0,1692
12	10	Euc	-	-	-	CFR	Recall	0,0706
-	80	-	-	-	-	SPA	F1	0,0981
-	80	-	-	-	-	SPA	Precision	0,1848
-	70	-	-	-	-	SPA	Recall	0,0849
11	80	Cos	7	3	-	CFR-CFG	F1	0,0767
11	80	Cos	7	3	-	CFR-CFG	Precision	0,1748
12	10	Euc	10	0	-	CFR-CFG	Recall	0,0706
12	80	Cos	9	-	1	CFR-SPA	F1	0,1032
12	80	Cos	9	-	1	CFR-SPA	Precision	0,2006
13	70	Cos	9	-	1	CFR-SPA	Recall	0,0878
10	80	Jac	-	0	10	CFG-SPA	F1	0,0984
10	80	Jac	-	0	10	CFG-SPA	Precision	0,1848
7	20	Cos	-	3	7	CFG-SPA	Recall	0,0852
13	80	Pea	8	1	1	CFR-CFG-SPA	F1	0,1033
13	80	Cos	4	0	6	CFR-CFG-SPA	Precision	0,2007
13	70	Cos	9	0	1	CFR-CFG-SPA	Recall	0,0874

when producing recommendation. We evaluate every iteration with Precision, Recall, and F1 score to determine the recommendation quality.

After we get the result of algorithm CFR, CFG, and SPA, then we hybrid those 3 algorithms to compare it with the singular algorithms. CFR-CFG, CFR-SPA, CFG-SPA, and then at the end we hybrid CFR-CFG-SPA to see whether the hypothesis are true, that by combining broader attribute, will increase the quality of the recommendation.

At the end of the experiment, we count how many times certain similarity could produce the best result to determine which similarity are the best.

3.1. Data Preparation

First, we'll filter the user by the number of ratings they've given, with a minimum of 30 ratings, and then we'll separate the dataset based on the order of consumption. We will allocate 70% of the dataset for data Training and 30% for data Testing. Due to time and resource constraints, this experiment will only calculate a data sample with a confidence level of 95% and an error rate of 5%. According to Wilmoth and Krejcie, et al,[32, 33] the appropriate sample size based on the total number of 6040 users is 361, but for this study, we will randomly select 400 users as samples.

After data partitioning, we do data pre-processing to get the necessary information for CFR and CFG algorithms. For CFR, we were required to use the rating information contained in the dataset, unmodified. But for CFG, we calculate the user's genre preference score by filtering each genre of the films viewed by the user, multiplying each film by its rating, and dividing each genre by the total number of films viewed. Then, we obtain the genre preference score of the user.

Similarity equations are utilized in the CFR and CFG algorithms to determine the neighbourhood of similar users. There are numerous methods for determining user similarity, including Pearson correlation [34, 35], and Cosine similarity [3, 36], thus we will do research to develop a CF algorithm employing these equations.

Choi, et al are comparing Pearson, Euclidean, and Cosine similarities to see which similarity equation yields the best recommendation.[16] In this experiment, we will add Jaccard similarity due to its simplicity to determine if it produces better recommendations than previously described similarities. Conceptually, Figure 1 illustrates the perspective of each similarity equation.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.2. Calculation

We obtain the N most similar users based on each similarity equation. Then, we multiply the correspondence similarity score by the given rating and divide the result by the sum of similarity scores to obtain the forecast rating for each film using below equation (1). Which R stands for ratings given by user, and S stands for similarity score.

$$Ru = \left(\sum_{u=1}^{n} R_u * S_u\right) / \left(\sum_{u=1}^{n} S_u\right) \quad (Eq.1)$$

In accordance with [30], SPA algorithms use support and confidence to determine which items are acceptable for recommendation; however, in this experiment, we will slightly alter the way projected score is generated.

Every occurrence of a candidate movie is multiplied by its rating by the corresponding user and the number of sets in the movie search pattern in order to reward movies with good ratings and the likelihood that the pattern occurred. This sum is then summed and divided by the number of times that movie was consumed.

3.3. Producing Recommendation List

After obtaining the predicted rating for each algorithm, we first normalize the score before combining them using a weighted approach. Using equation (2), we combine every feasible combination, namely CFR with CFG, CFR with SPA, and CFG with SPA. Symbol α represents a decimal number between 0 and 1 with steps of 0,1.

$$P(a) = (\alpha \ x \ Eq1) + ((1 - \alpha) \ x \ Eq2) \ (Eq.2)$$

We normalize and aggregate the CFR-CFG-SPA score using equation (3), and then assess each technique to see which produces the best suggestions. Symbol α, β, and γ are decimal numbers between 0 and 1 with steps of 0.1, where α+β+γ=1.

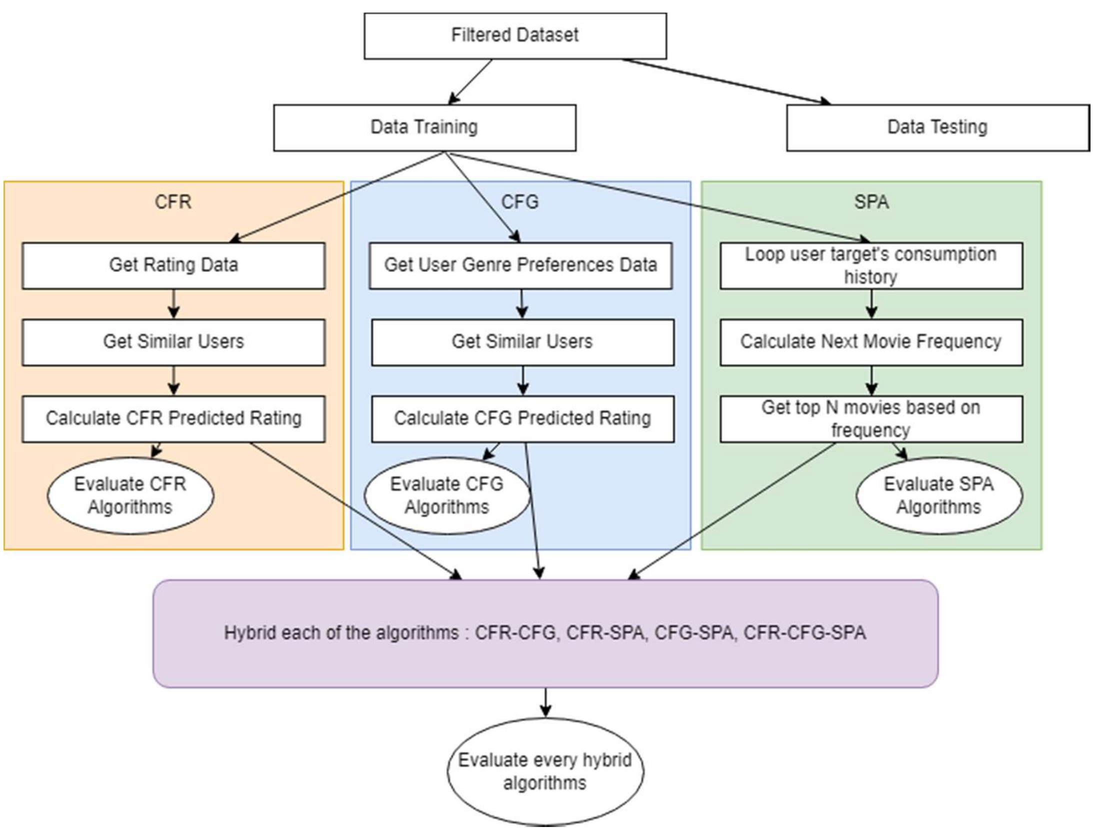

Figure 2: Hybrid Algorithms Overview

$$P(a) = (\alpha\ x\ CFR) + (\beta\ x\ CFG) + (\gamma\ x\ SPA) \qquad (Eq. 3)$$

Figure 2 provides an overview of hybrid algorithms.

4. EXPERIMENT RESULT

We examine seven algorithms and compare them to one another. In terms of precision, recall, and f1 score, the objective is to determine if hybrid algorithms produce better recommendations than single algorithms.

4.1. Experiment 1

The goal of this experiment is to find out which parameter produces the best recommendation for algorithms CFR, CFG, and SPA. The parameters being measured are k-user, which indicates how many similar users are used to compute predicted ratings for each algorithm, k-list, which indicates how many items should be recommended for each method, and similarity function. We conduct experiments with 7, 8, 9, 10, 11, 12, and 13 users for k-user. For k-list, we do experiments with 10, 20, 30, 40, 50, 60, 70, and 80 recommended items. We experiment with Pearson, Cosine, Euclidean, and Jaccard similarity measures.

Table 1 identifies the optimal parameter for producing the best outcomes for each algorithm and evaluation. Observing the findings reveals that precision and f1 evaluation results are optimally generated with the k-user ad k-list parameter set to its maximum or near-maximum value. On the other hand, the stories are quite the opposite for recall scores. The majority of the algorithm's optimal outcomes are produced using cosine similarity.

4.1.1. CFG Algorithm

We find that a combination of 13 k-users, 80 k-lists, and a cosine similarity function yields the best f1 results, of 0,0662, when employing CFG algorithms. Six of the top ten f1 outcomes employ cosine similarity, whilst four employ pearson similarity. With the majority of it utilizing 80 k-list and 9-13 k-users. 13 k-user, 80 k-list, and cosine similarity are used to get the highest precision, which is 0.1233, with 13 k-user, 80 k-list, and cosine similarity. Five of the top ten precision findings use cosine similarity, while the other five use pearson similarity. With each of them employing 80 k-lists and a value between 9 and 13 for the k-users option. Using 12 k-user, 10 k-list, and cosine similarity yields the best recall results of 0.0667. Six of the top ten recall score results use cosine similarity, three

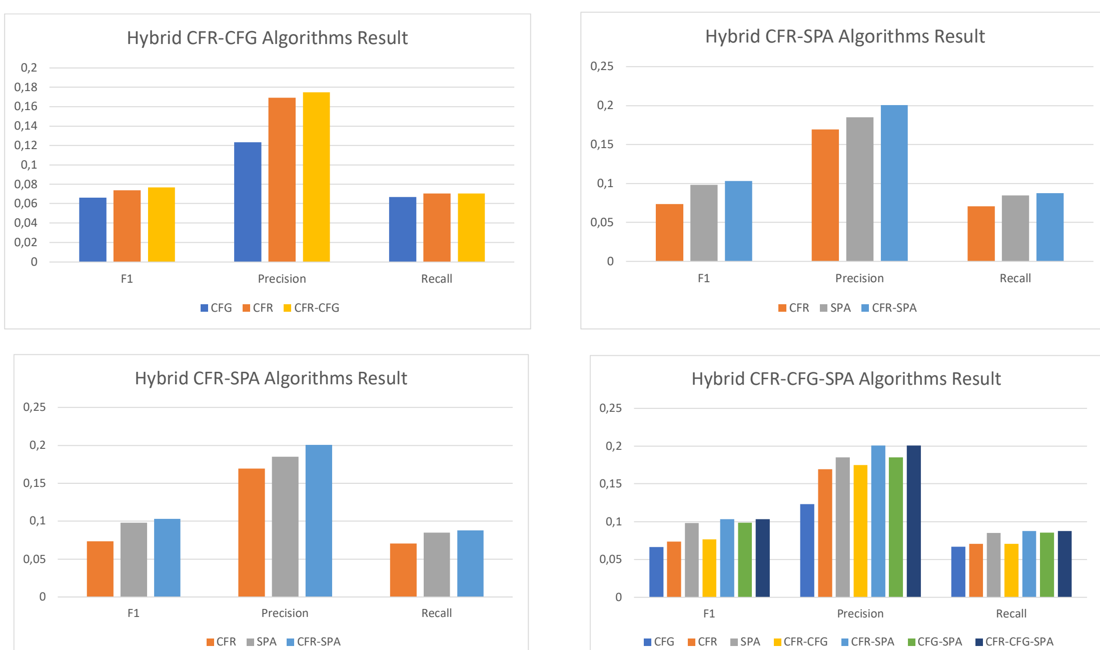

Figure 3: Result Comparison Between Algorithms

use Pearson similarity, and the remaining four use Euclidean similarity.

4.1.2. CFR Algorithm

Using 11 k-user, 80 k-list, and jaccard similarity, CFR algorithms produce the best f1 results, which is 0.0736, by utilizing jaccard similarity. Five of the top ten f1 outcomes employ jaccard similarity while the other five use cosine similarity, with each result using 80 k-lists and a k-user parameter ranging from 7 to 13. For best precision results (0.1692), 12 k-users, 80 k-lists, and cosine similarity are used; of the top 10 results, 6 utilize cosine similarity while the remaining 4 use pearson similarity. All of the top 10 results utilize 80 k-list and vary from 7 to 13 k-user. The optimal recall score is 0.0706 while utilizing 12 k-user, 10 k-list, and euclidean similarity. All of the top 10 recall results employ euclidean similarity, the range of k-list is between 10 and 30, and the range of k-user is 9 to 13.CFR-CFG-SPA and CFR-SPA have distinguishable weight and k-user parameters, but same similarity equations and k-list parameters, for their precision scores. The precision scores are not significantly different but still prefer the CFR-CFG-SPA algorithms.

4.1.3. SPA Algorithm

We did not do any studies on k-user and similarity functions as SPA techniques do not employ similarity equations. Consequently, we only experiment with the k-list parameter. F1 score and Precision score highest results, which are 0.0981 and 0.1848, respectively, use the same k-list parameter, 80, whereas the best recall score, 0.0849, uses 70 k-list.

4.1.4. Hybrid CFR-CFG Algorithm

The best recall score produced by the CFR-CFG-SPA method has the same weight, k-user parameter, and k-list parameter as the CFR-SPA algorithm; even for the similarity, cosine similarity, is the same. This demonstrated that the CFG algorithm did not contribute to creating better recommendations based on the recall score of the CFR-CFG-SPA method.

The CFR-CFG method outperforms the CFG and CFR algorithms by 13,63% and 4,04% for the F1 score, 29,46% and 3.2% for the precision score, and 5.59 and 0.01% for the recall score, respectively, as seen in the graph in Figure 3's upper left corner.

4.1.5. Hybrid CFR-SPA Algorithm

As displayed in the top right-hand corner of Figure 3, the CFR-SPA algorithm achieves superior outcomes to the CFR and SPA algorithms by 28,71% and 4,92% for the F1 score, 15,65% and 7,85% for the precision score, and 19,59% and 3,33% for the recall score, respectively.

4.1.6. Hybrid CFG-SPA Algorithm

The CFG-SPA algorithm outperforms the CFG and SPA algorithms by 32.71% and 0.28% for the F1 score, 33.29% and -0.01% for the precision score, and 21.75% and 0.36% for the recall score, as seen in the lower left chart of Figure 3.

4.1.7. Hybrid CFR-CFG-SPA Algorithm

The CFR-CFG-SPA hybrid algorithm produces the most accurate recommendations by 25,77%, 0,09%, and 4,74% compared to algorithm CFR-CFG, CFR-SPA, and CFG-SPA, based on the F1 score, also the best performance by 12,91%, 0,06%, and 7,91% based on precision measurement. However, the CFR-SPA method produced a slightly higher recall score than the CFR-CFG-SPA algorithm. Observing the weight parameter for computation of precision and recall, algorithm CFG does not contribute to the CFR-CFG-SPA method's performance. On the other hand, the CFG algorithm improves the quality of the recommendation based on the f1 score.

4.2. Experiment 2

In this experiment, we will examine the characteristics of each algorithm's recommendations. In this instance, we will use user with id 6016. The target audience has viewed a total of 328 films, with drama being the preferred genre, followed by thriller, romance, western, comedy, and crime.

Surprisingly, the CFG algorithm as shown on Table 2 recommends drama and comedy films less frequently than action films, even though these are the genres that the user prefers. The CFR algorithm, on the other hand as shown on Table 3, recommends a greater variety of genres, with the majority of them being dramas and crime films, while producing a higher f1 score than the CFG algorithm. This algorithm provides the best f1 score when compared

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

to CFG and CFR. The SPA algorithm shown in Table 4, produced a majority of horror, thriller, and drama films, which are the genres that users consume the least.

CFR-CFG algorithms as shown on Table 5, give more diverse movie recommendations. There are film noir, action, and science fiction films available. Even though the genres are broad, and the user hasn't seen many movies in these genres before, this hybrid algorithm makes better movie suggestions than algorithms that work on their own. On the other hand, the CFR-SPA algorithm as shown on Table 6, generates a familiar genre comprised mostly of action, crime, and drama movie recommendations, as well as film-noir and fantasy. On the CFG-SPA algorithm, however, as shown on Table 7, produced similar list of proposed films is compared to the SPA method, as the CFG algorithm had no effect on the CFG-SPA algorithm's precision. Compared to other algorithms, the CFR-CFG-SPA algorithm is the most diverse and produces the most accurate movie suggestion list as shown on Table 8. It possesses the qualities of each algorithm and generates the most original movie suggestion compared to the others.

5. CONCLUSION

After doing the experiment, we can infer that by including genre as one of the attribute the CFR-CFG-SPA hybrid algorithms produce better recommendations than the other algorithms tested. On the other hand, CFR-SPA algorithms produce the highest recall score. Majority of the algorithms generate superior recommendations by utilizing cosine similarity. Also in the majority of algorithms, the optimal range for the number of similar users utilized to get a predicted rating is between 10 and 13, but this aspect requires further study.

This experiment concludes that Jaccard similarity are not the best similarity to produce recommendation. Cosine similarity is the similarity that produces the best outcomes the majority of the time for f1, precision, and recall in all algorithms with the exception of the f1 score on CFR and recall score on CFR. Maximum k-lists should be used for f1 score and precision score on all algorithms, but minimum k-lists should be used for recall score because recall score is intended to reward false negative outcomes.

ACKNOWLEDGMENT

The author wishes to express gratitude to everyone who contributed to the creation of this paper. By being a kind discussion partner who occasionally offered the authors encouragement. The author would also like to thank Mr. Gregorius Natanael and Mr. Rusnandi Fikri for their inspiration in conducting the experiment for this research.

REFERENCES:

[1] C.L. Huang and W.L. Huang, "Handling sequential pattern decay: Developing a two-stage collaborative recommender system.," *Electronic Commerce Research and Applications*. vol. 8, no. 3, pp. 117–129, 2009.
[2] Z. Wang, X. Yu, N. Feng, and Z. Wang, "An improved collaborative movie recommendation system using computational intelligence.," *Journal of Visual Languages and Computing*. vol. 25, no. 6, pp. 667–675, 2014.
[3] K.C. Lee and S. Kwon, "Online shopping recommendation mechanism and its influence on consumer decisions and behaviors: A causal map approach.," *Expert Systems with Applications*. vol. 35, no. 4, pp. 1567–1574, 2008.
[4] C.C. Aggarwal, *Recommender Systems: The Textbook*. *Springer International Publishing*, Cham, 2016.
[5] P. Maes, R.H. Guttman, and A.G. Moukas, "Agents that buy and sell.," *Communications of the ACM*. vol. 42, no. 3, p. 81, 1999.
[6] J. ben Schafer, J.A. Konstan, and J. Riedl, "E-Commerce Recommendation Applications.," *Data Mining and Knowledge Discovery 2001 5:1*. vol. 5, no. 1, pp. 115–153, 2001.
[7] R. Burke, "Hybrid Recommender Systems: Survey and Experiments.," *User Modeling and User-Adapted Interaction 2002 12:4*. vol. 12, no. 4, pp. 331–370, 2002.
[8] A. Ansari, S. Essegaier, and R. Kohli, "Internet Recommendation Systems.," *Journal of Marketing Research*. vol. 37, no. 3, pp. 363–375, 2000.
[9] H.J. Lee and J.K. Lee, "An effective customization procedure with configurable standard models.," *Decision Support Systems*. vol. 41, no. 1, pp. 262–278, 2005.
[10] P. Pu and Li Chen, "A User-Centric Evaluation Framework of Recommender Systems.," *Proceedings of the fifth ACM confer- ence on Recommender Systems (RecSys'11), ACM, New York, NY, USA;* pp. 57–164, 2011.

[11] P. B.Thorat, R.M. Goudar, and S. Barve, "Survey on Collaborative Filtering, Content-based Filtering and Hybrid Recommendation System.," *International Journal of Computer Applications*. vol. 110, no. 4, pp. 31–36, 2015.

[12] I. MacKenzie, C. Meyer, and S. Noble, "How retailers can keep up with consumers," https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers.

[13] Y.Y. Shih and D.R. Liu, "Hybrid recommendation approaches: Collaborative filtering via valuable content information.," *Proceedings of the Annual Hawaii International Conference on System Sciences*. p. 217, 2005.

[14] J. Lampel and H. Mintzberg, "Customizing Customization.," *MIT Sloan Management Review*. p. 1996.

[15] B.P.S. Murthi and S. Sarkar, "The Role of the Management Sciences in Research on Personalization.," *http://dx.doi.org/10.1287/mnsc.49.10.1344.17313*. vol. 49, no. 10, pp. 1344–1362, 2003.

[16] K. Choi, D. Yoo, G. Kim, and Y. Suh, "A hybrid online-product recommendation system: Combining implicit rating-based collaborative filtering and sequential pattern analysis.," *Electronic Commerce Research and Applications*. vol. 11, no. 4, pp. 309–317, 2012.

[17] C.N. Sunilkumar and Goyani Mahesh, "A review of movie recommendation system: Limitations, Survey and Challenges.," *ELCVIA Electronic Letters on Computer Vision and Image Analysis*. vol. 19, no. 3, pp. 18–37, 2020.

[18] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews.," *Proceedings of the 1994 ACM conference on Computer supported cooperative work - CSCW '94*. p.

[19] D. Goldberg, D. Nichols, B.M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry.," *Communications of the ACM*. vol. 35, no. 12, pp. 61–70, 1992.

[20] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Analysis of Recommendation Algorithms for E-Commerce.," In: *Proceedings of the 2nd ACM Conference on Electronic Commerce*. pp. 158–167. *Association for Computing Machinery*, New York, NY, USA (2000).

[21] U. Shardanand and P. Maes, "Social Information Filtering: Algorithms for Automating 'Word of Mouth,.'" In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. pp. 210–217. *ACM Press/Addison-Wesley Publishing Co.*, USA (1995).

[22] J. Rucker and M.J. Polanco, "Siteseer.," *Communications of the ACM*. vol. 40, no. 3, pp. 73–75, 1997.

[23] J. Delgado and Ishii, "Memory-Based Weighted-Majority Prediction for Recommender Systems.," In: *SIGIR Workshop Recomm. Syst. Citeseer*. pp. 85 (1999).

[24] L. Si and R. Jin, "Flexible Mixture Model for Collaborative Filtering.," p. 2003.

[25] Y. Koren, "Collaborative filtering with temporal dynamics.," *Communications of the ACM*. vol. 53, no. 4, pp. 89–97, 2010.

[26] M. Pazzani, D. Billsus, R.S. Michalski, and J. Wnek, "Learning and Revising User Profiles: The Identification of Interesting Web Sites.," *Machine Learning 1997 27:3*. vol. 27, no. 3, pp. 313–331, 1997.

[27] G. Geetha, M. Safa, C. Fancy, and D. Saranya, "A Hybrid Approach using Collaborative filtering and Content based Filtering for Recommender System.," In: *Journal of Physics: Conference Series*. pp. 012101. *IOP Publishing* (2018).

[28] D.R. Liu, C.H. Lai, and W.J. Lee, "A hybrid of sequential rules and collaborative filtering for product recommendation.," *Information Sciences*. vol. 179, no. 20, pp. 3505–3519, 2009.

[29] R. Devooght and H. Bersini, "Long and short-Term recommendations with recurrent neural networks.," *UMAP 2017 - Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*. pp. 13–21, 2017.

[30] R. Agrawal and R. Srikant, *Mining Sequential Patterns*., 1995.

[31] F.M. Harper and J.A. Konstan, "The MovieLens Datasets.," *ACM Transactions on Interactive Intelligent Systems*. vol. 5, no. 4, pp. 1–19, 2016.

[32] G.H. Wilmoth, "Handbook in research and evaluation, Second Edition Stephen Isaac and William B. Michael San Diego, CA: EDITS Pubs., 1981.," *https://doi.org/10.1177/105960118200700111*. vol. 7, no. 1, pp. 124–126, 1981.

[33] R. V. Krejcie and D.W. Morgan, "Determining Sample Size for Research Activities.," *http://dx.doi.org/10.1177/001316447003000308*. vol. 30, no. 3, pp. 607–610, 1970.

[34] A. Albadvi and M. Shahbazi, "A hybrid recommendation technique based on product category attributes.," *Expert Systems with Applications*. vol. 36, no. 9, pp. 11480–11488, 2009.

[35] M. Kwak and D.S. Cho, "Collaborative filtering with automatic rating for recommendation.," *IEEE International Symposium on Industrial Electronics*. vol. 1, pp. 625–628, 2001.

[36] P. Aggarwal, V. Tomar, and A. Kathuria, "Comparing Content Based and Collaborative Filtering in Recommender Systems.," *International Journal of New Technology and Research*. vol. 3, no. 4, p. 2017.

Journal of Theoretical and Applied Information Technology
15th August 2023. Vol.101. No 15

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

APPENDIX

Table 2: Recommendation List Produced by CFG Algorithm.

	movie	title	genres
0	movie_1214	Alien (1979)	Action|Horror|Sci-Fi|Thriller
1	movie_260	Star Wars: Episode IV - A New Hope (1977)	Action|Adventure|Fantasy|Sci-Fi
2	movie_1196	Star Wars: Episode V - The Empire Strikes Back...	Action|Adventure|Drama|Sci-Fi|War
3	movie_1210	Star Wars: Episode VI - Return of the Jedi (1983)	Action|Adventure|Romance|Sci-Fi|War
4	movie_2762	Sixth Sense, The (1999)	Thriller
5	movie_1240	Terminator, The (1984)	Action|Sci-Fi|Thriller
6	movie_924	2001: A Space Odyssey (1968)	Drama|Mystery|Sci-Fi|Thriller
7	movie_1270	Back to the Future (1985)	Comedy|Sci-Fi
8	movie_1200	Aliens (1986)	Action|Sci-Fi|Thriller|War
9	movie_1198	Raiders of the Lost Ark (1981)	Action|Adventure

Table 3: Recommendation List Produced by CFR Algorithm.

	movie	title	genres
0	movie_858	Godfather, The (1972)	Action|Crime|Drama
1	movie_1387	Jaws (1975)	Action|Horror
2	movie_593	Silence of the Lambs, The (1991)	Drama|Thriller
3	movie_296	Pulp Fiction (1994)	Crime|Drama
4	movie_50	Usual Suspects, The (1995)	Crime|Thriller
5	movie_260	Star Wars: Episode IV - A New Hope (1977)	Action|Adventure|Fantasy|Sci-Fi
6	movie_1198	Raiders of the Lost Ark (1981)	Action|Adventure
7	movie_608	Fargo (1996)	Crime|Drama|Thriller
8	movie_1221	Godfather: Part II, The (1974)	Action|Crime|Drama
9	movie_111	Taxi Driver (1976)	Drama|Thriller

Table 4: Recommendation List Produced by SPA Algorithm.

	movie	title	genres
0	movie_1278	Young Frankenstein (1974)	Comedy|Horror
1	movie_1219	Psycho (1960)	Horror|Thriller
2	movie_3671	Blazing Saddles (1974)	Comedy|Western
3	movie_1080	Monty Python's Life of Brian (1979)	Comedy
4	movie_1231	Right Stuff, The (1983)	Drama
5	movie_924	2001: A Space Odyssey (1968)	Drama|Mystery|Sci-Fi|Thriller
6	movie_1304	Butch Cassidy and the Sundance Kid (1969)	Action|Comedy|Western
7	movie_2067	Doctor Zhivago (1965)	Drama|Romance|War
8	movie_1387	Jaws (1975)	Action|Horror
9	movie_1258	Shining, The (1980)	Horror

Table 5: Recommendation List Produced by CFR-CFG Algorithm.

	movie	title	genres
0	movie_260	Star Wars: Episode IV - A New Hope (1977)	Action|Adventure|Fantasy|Sci-Fi
1	movie_1240	Terminator, The (1984)	Action|Sci-Fi|Thriller
2	movie_1214	Alien (1979)	Action|Horror|Sci-Fi|Thriller
3	movie_1198	Raiders of the Lost Ark (1981)	Action|Adventure
4	movie_1210	Star Wars: Episode VI - Return of the Jedi (1983)	Action|Adventure|Romance|Sci-Fi|War
5	movie_1617	L.A. Confidential (1997)	Crime|Film-Noir|Mystery|Thriller
6	movie_1196	Star Wars: Episode V - The Empire Strikes Back...	Action|Adventure|Drama|Sci-Fi|War
7	movie_1193	One Flew Over the Cuckoo's Nest (1975)	Drama
8	movie_1270	Back to the Future (1985)	Comedy|Sci-Fi
9	movie_593	Silence of the Lambs, The (1991)	Drama|Thriller
7	movie_2067	Doctor Zhivago (1965)	Drama|Romance|War
8	movie_1387	Jaws (1975)	Action|Horror
9	movie_1258	Shining, The (1980)	Horror