An Efficient Framework of Hybrid Recommendation System based on Multi Mode

Tessy Badriyah
Politeknik Elektronika Negeri Surabaya, INDONESIA
tessy@pens.ac.id

Yunaz Gilang Ramadhan
Politeknik Elektronika Negeri Surabaya, INDONESIA
yunaz.gilang@gmail.com

Iwan Syarif
Politeknik Elektronika Negeri Surabaya, INDONESIA
iwanarif@pens.ac.id

***Abstract*— Recommendation systems have been widely applied in many areas, such as E-commerce, and so on. However, in some complex systems such as missed sparse data, it will be increasingly difficult to build a model for user recommendations. In this research we develop a recommendation system on E-Commerce. This system will be able to adapt and provide the best recommendations for each user dynamically even in sparse environment. The system will be created in a web-based application to display the product recommendations to users. The recommendation system developed is expected to be able to solve cold-start problem when there is no other relevant data to be recommended for the new added product and also the sparsity problem. To overcome this problem, the system will implement multi-mode algorithm that uses more than one search algorithm for the closest characteristics in the recommendation system and can choose one of the best algorithms to use in accordance with the existing data and hybrid-filtering that can use a combination of Collaborative Filtering is to make recommendations based on information equations between users and Content-Based Filtering is to make recommendations based on information representation of a content. Thus the system will be able to provide product recommendations on any state of data on E-Commerce.**

Keywords*— *Recommender system, sparsity, cold-start, hybrid filtering

I. Introduction

Technology is growing faster. Today we are entering a digital age where many people who have access to the internet and information are important in their daily lives. Realizing this, business people began to adapt. In the past, business people carried out their business using conventional methods, however, today many business people use information technology to market their products or services. One of the uses of information technology for business is E-Commerce. E-Commerce itself is currently growing very rapidly because of the ease of transaction offered. E-Commerce systems are becoming increasingly complex because more products or services are offered, and more users are available. The complexity of information contained in this E-Commerce system will be very difficult to manage. Users will also find it difficult to get the product or service he wants. Sometimes many prospective buyers do not know what items to buy, they will try to find information about what products are offered. Problems arise because a lot of information must be processed, imagine just to make a purchase, the user must do a lot of page switching. Even using the search engine features is not the solution, because search results will always be the same for all users, even though users only want to buy the product that they think is the best. The recommendation system is the right solution to solve the problem.

However, there is no method in the recommendation system that is good to use at any time. There are several conditions in which a recommendation system method is good to use than other recommended system methods. The reality is not all E-Commerce has the same characteristics and the possibility with the development of business carried out by an E-Commerce can also experience changes in characteristics. This will make a good recommendation system used before it can turn out to be bad. For this reason, it is necessary to develop a recommendation system that is able to adapt according to the characteristics of information contained in E-Commerce. The flexibility of the recommendation system will produce recommendations that are always good for each situation. Research on the construction of a flexible recommendation system is not only applicable in the case of product recommendations in E-Commerce but can also be useful in developing a system of recommendations for other cases.

II. The Proposed Method

In this study, we built a separate multi-mode method which is able to accept various kinds of data input (rating, order, visit, & search). The proposed method called Multi-Mode Hybrid Filtering (MMHF). Hybrid filtering method in this research can formulate recommendations based on the weighting average of the recommendations using both Collaborative and Content-Based Filtering. The weighting value is obtained from multiplying the calculation results from Association Retrieval Correlation (ARC) with the similarity value calculated by Content based filtering for each product. The following will explain the system design of the research that we did and the part that underlies the method we propose, namely Association Retrieval Correlation and Content Based Filtering by Text -Mining on item's information.

A. A. System Design

The proposed system consists of data extraction features which are the pre-processing stage to obtain data and store important features on each data (Fig 1., number 1).

978-1-7281-9530-8/20/$31.00 ©2020 IEEE

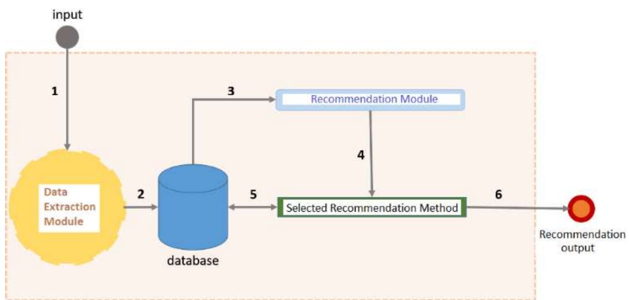

Figure 1. System Design

The data obtained is then stored in a database (shown in number 2). When there is a request to compile recommendations, the system will access information from the database. This information will be processed by a module that determines recommendations that will use the Association Retrieval Correlation (ARC) as Collaborative Filtering and will combine the results among products using the Content based Filtering method based on the product features, The product features in this case are a combination of text-mining and TF-IDF calculations (indicated by number 3) . The process will produce a recommendation module that automatically selects the data to be processed, and other Hybrid Filtering parameters to process data and arrange recommendations (shown in number 4). The selected recommendation mode will then formulate recommendations (shown in number 5). The output in the form of recommendations will then be obtained by the user (shown in number 6).

III. Tracking User Behaviour on Multi-Mode Hybrid Filtering

Tracking user behavior aims to provide dynamic recommendations even though user information is not obtained from rating or order data, even by applying this module, users who are not logged in are still able to get recommendations according to their habits. Currently in our design, there are two user habits that are recorded. Namely the history of search and history of user visit to the product

User history is valuable information for the recommendation system. But not all information is always relevant. For example, for searches, we can find out the tendency of current users to prefer products like enough from the latest search history, the old history is no longer relevant. To achieve what is needed, we apply FIFO (First In First Out) to the variable that stores the user's search history

The following diagram explains the workflow from the tracking process of user behavior in MMHF.

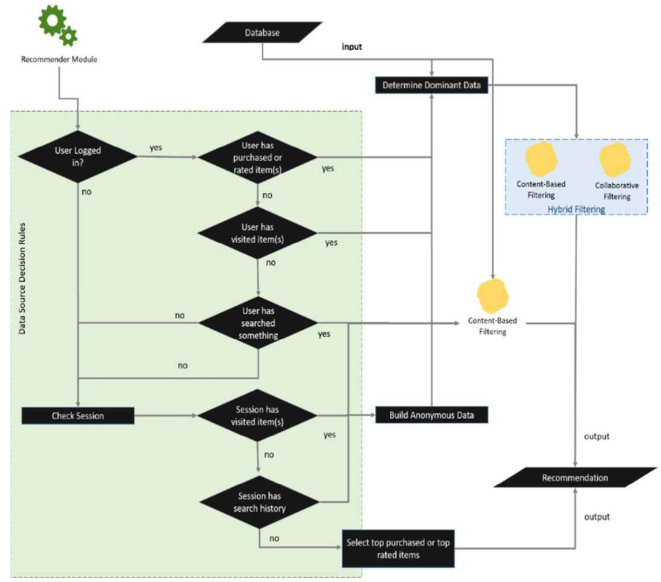

Figure 2. Workflow of Multi Mode Hybrid Filtering (MMHF) Recommendation System

The recommendation system designed has priority in selecting input data and methods used.

The following is the priority sequence of input data starting from the lowest to the highest:

a. Search History

Retrieving data entered from the user's search history, will only be used when other data is not available. The filtering method used when using this data is Content-Base Filtering. Search for the most relevant products with users based on their search history.

b. Visit History

Retrieving data entered from the user's visit history, will only be used when rating data and orders are not available. The filtering method used is Hybrid Filtering, with comparison of order data or other user rating data.

c. Rating

Retrieve data entered from the user rating history of the product. Used according to the tendency of users who prefer to give a rating on the product. The filtering method used is Hybrid Filtering with a comparison of other user rating data

d. Order

Retrieve data entered from the user's purchase history. Used according to the tendency of users who prefer to buy products. The filtering method used is Hybrid Filtering with comparison of other user's order data.

The above information will be processed to find recommendations and after that the results of recommendations can be evaluated

IV. Result and Discussion

In this study, several stages of testing were carried out to prove the superiority of the proposed hybrid filtering method and see whether the recommendation system that had been designed could work well. The reference for assessing how well the results obtained is determined by the performance measurement used in this study

A. Performance Measurement

There are several ways to determine the performance of a recommendation system. Following are the methods used to calculate system recommendation performance in this study:

1. **Precision :**
   Is the ratio of the number of recommendations relevant to the total number of recommendations.

$$Precision(p) = \frac{tp}{(tp+fp)} \quad (1)$$

Where:

- tp is the number of products found and relevant
- fp is the number of products found but not relevant

2. **Recall :**
   Is the ratio of the number of recommendations relevant to the total relevant data.

$$Recall(r) = \frac{tp}{(tp+fn)} \quad (2)$$

Where:

- tp is the number of products found and relevant
- fn is the number of products not found but relevant

3. **Fallout :**
   Is the ratio of the number of recommendations given irrelevant to the total irrelevant data.

$$Fallout(f) = \frac{fp}{(fp+tn)} \quad (3)$$

Where:

- fp is the number of products found but not relevant
- tn is the number of irrelevant products

4. **missRate**
   is the ratio of the number of relevant items but is not recommended with the total relevance of the relevant items.

$$missRate(m) = \frac{fn}{(tp+fn)} \quad (4)$$

Where:

- fn is the number of products that are relevant but not recommended
- tp is the recommended and relevant number of products

5. **F-measure :**
   An evaluation calculation in an information retrieval that combines recall and precision. The recall value and precision in a situation can have different weights. The measure that displays reciprocity between recall and precision is a F-measure which is the mean harmonic weight of recall and precision [13]

$$F-measure(fm) = \frac{2xpxr}{p+r} \quad (5)$$

Where:

- p is the value of precision
- r is the recall value

B. Testing Scenario

Comparative data scenarios are recommendations based on scenarios compared to a user's relevant data at that time to measure the performance recommendations given. This scenario is divided into two types, namely strict and non-strict. Strict mode means that the recommendation data will only be compared with the original data from the user. Missing recommendations get data sources from user rating data. Then the recommendations will be compared to products that have been rated by that user only. While the second scenario in the form of non-strict mode means that when recommendations get data sources from rating data, the results of the recommendations will be compared to products from the top 5 categories that are rated best by users.

Before discussing the results of the method, we will display data visualization to split two data entries from user orders and ratings:

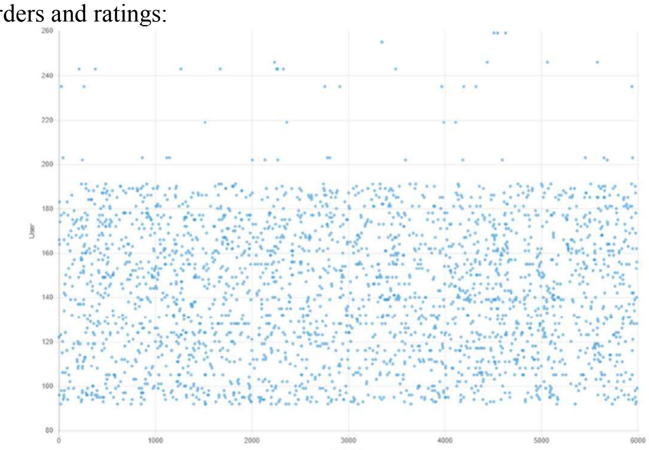

Figure 3. Distribution of order data

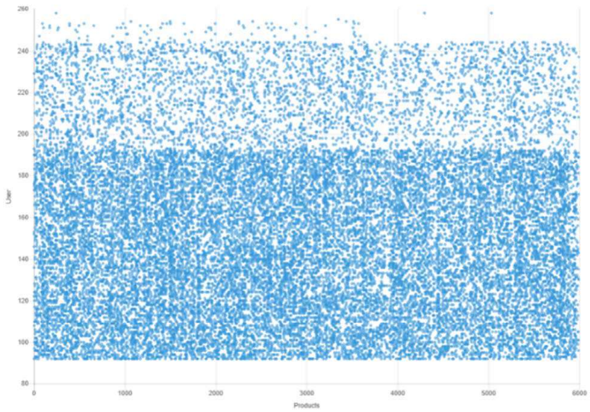

Figure 4. Distribution of rating data

From the data distribution in Figure 3 and 4, it can be seen that the order data is more sparse. This more sparse condition will certainly affect the value of confidence. The more tightly the value of confidence will be higher too, and the easier it is to recommend other products for users

C. *Performance Analysis*

In this section, an experiment is conducted to compare the performance between the hybrid method and the collaborative filtering method. The Hybrid Filtering method used is a combination of ARC and Content-Base Filtering methods

The comparison scenarios are divided into two types, namely strict and non-strict. Strict mode means that the recommendation data will only be compared with the original data from the user, for example the recommendation gets the data source from the user rating data, so the recommendation results are only compared to products that have been rated by the user. While the second scenario is a non-strict mode, meaning that when the recommendation gets the data source from the rating data, the recommendation results will be compared with the products from the top 5 categories that are rated the best by the user.

The hybrid filtering module used has adjustable parameters so that it can provide flexibility to determine the number of recommendations. For example, with the value of LIMIT_COLLABORATIVE = 100 and LIMIT_CONTENT_BASE = 10, the maximum recommendation obtained is 100 x 10 = 1000 products (this number will later decrease because the system will ignore the same recommended product).

In the experiment, we use three methods, namely collaborative filtering method, hybridA with collaborative weight 100 and content-base 10 weight, while hybridB with collaborative weight 50 and content-base weight 10. All of these scenarios are run on product visit history data and rating data given by the customer to the product.

(i). Using product visit data:

Experiments on product visit history data using new users who have only visited one product. And will be compared the results of recommendations from which algorithm can recommend more products.

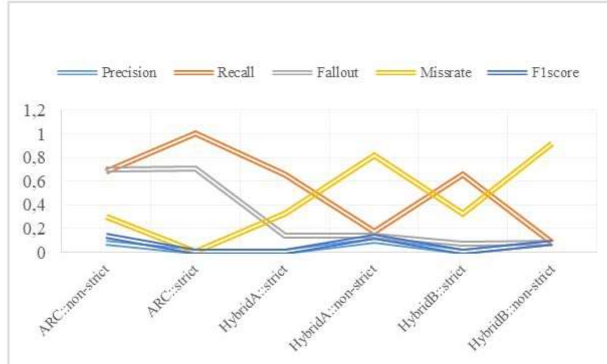

Figure 5. User Visit Evaluation

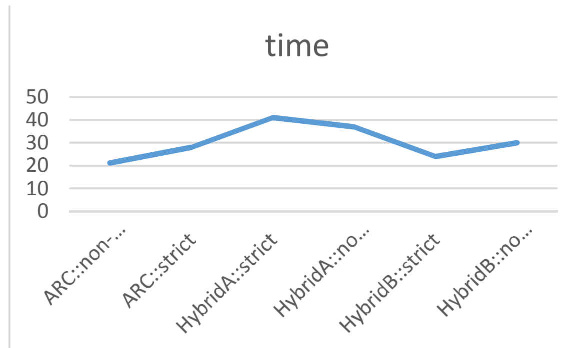

Figure 6. Computation time on the User Visit

From Fig 5 and Fig 6 , it can be seen that collaborative performance is superior to Hybrid in fallout, recall, missrate and time execution. But the Hybrid method is superior to the precision and f-1-score. It can be concluded that the more collaborative weights on hybrid filtering, the longer the execution time will be.

(ii). Using product rating data:

Meanwhile, the evaluation of the rating data test will compare which algorithm is the most reliable in handling less sparse data

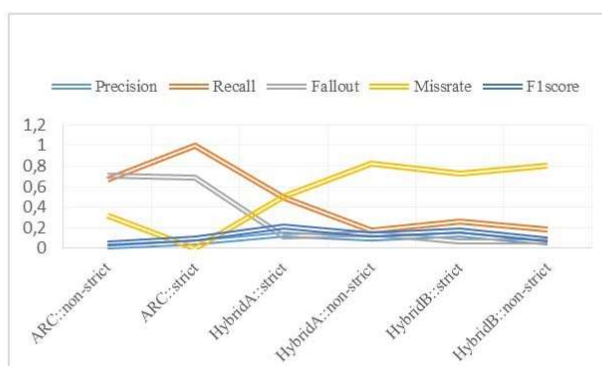

Figure 7. Product Rating Data Evaluation

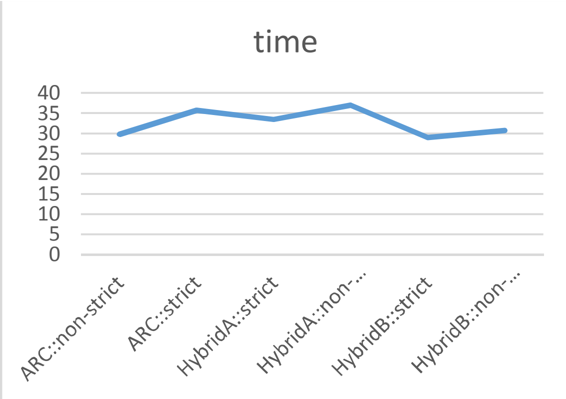

Figure 8. Computation time on Product Rating Data

As previous experiments, collaborative has good recall values, fallout, and miss rate. But collaborative is not consistent in strict and non-strict modes. While hybrids are more consistent with results. Besides that, it was found, the collaborative algorithm is very wasteful of time for large amounts of data (such as rating data)

V. Conclusion

This study proposes an efficient Hybrid Recommendation System based on Multi Mode. The proposed method called Multi Mode Hybrid Filtering (MMHF) is based on user behavior tracking, which is about history of searching and history of user visits to products. In the experiment of the performance of features, the track module of user behavior has succeeded in efficiently monitoring user habits and formulating recommendations based on these data.

From the results of the experiment, it was found that the system can generate tags / features automatically through the text mining method and similarity calculations using TF-IDF smooth and with excellent content-based quality.

The track user behavior module has succeeded in overcoming the cold start problem in the system. To improve the performance of the feature, new information can be added which is monitored and integrated into the recommendation module. Besides that, the slow computation time is an obstacle that arises. Further research is needed on architectural design or system features in order to speed up the process of making recommendations.

References

[1] H. R. V. Paul Resnick, "Recommender System," *Communications of the ACM,* vol. 40, no. 3, pp. 56-58, 1997.

[2] V. Beal, "What is Recommender System?," www.webopedia.com, 2013. [Online]. Available: http://www.webopedia.com/TERM/R/recommender_systems.html. [Diakses 6 6 2017].

[3] M. C. J. H. Zhao Qiu, "Design of Multi-mode E-commerce Recommendation System," *Third International Symposium on Intelligent Information Technology and Security Informatics,* pp. 530-533, 2010.

[4] I. H. Witten, "Text Mining," University of Waikato, Hamilton, 2002.

[5] E. T. W. I. S. Tessy Badriyah, "A Hybrid Recommendation System for E-Commerce based on Product Description and User Profile," 2015.

[6] F. R. A. V. Tariq Mahmood, "Improving Recommendation Effectiveness by Adapting the Dialogue Strategy in Online Travel Planning," ECTRL Solutions, Trento, 2009.

[7] S. Robertson, "Understanding Inverse Document Frequency: On theoretical arguments for IDF," *Journal of Documentation,* vol. 60, no. 5, pp. 503-520.

[8] V. S. Prem Melville, "Recommender Systems," *Encyclopedia of Machine Learning,* pp. 829-838, 2010.

[9] "Recommender system," wikipedia.org, 31 December 2017. [Online]. Available: https://en.wikipedia.org/wiki/Recommender_systeHym. [Diakses 23 January 2018].

[10] M. T. W. L. Gunnar Schröder, "Setting Goals and Choosing Metrics for Recommender System Evaluations," European Social Fund and the Free State of Saxony, Dresden.

[11] C. W. M. X. X. G. YiBo Chen, "Solving the Sparsity Problem in Recommender Systems Using Association Retrieval," *Journal of Computers,* vol. 6, no. 9, pp. 1896-1902, 2011.

[12] F. O. A. H. J. B. Jesus Bobadilla, "A Collaborative Filtering Approach To Mitigate The New User Cold Start Problem," *Knowledge-Based Systems,* p. 225, 2012.

[13] Y. C. Y. J. M. S. Tan Yang, "A New Algorithm For Multi Mode Recommendation In Social Tagging Systems," 2012.

[14] J. T. X. Z. X. X. Mi Zhang, "Addressing Cold Start in Recommender Systems: A Semi-supervised Co-training Algorithm," SIGIR, 2014.

[15] A. M. Hridya Sobhanam, "Addressing cold start problem in recommender systems using association rules and clustering technique," dalam *International Conference on Computer Communication and Informatics*, Coimbatore, INDIA, 2013.