# Implementation of Weighted Parallel Hybrid Recommender Systems for e-Commerce in Indonesia

Mustika Aprilianti[1], Rahmad Mahendra[2], Indra Budi[3]
Faculty of Computer Science
Universitas Indonesia
mustikaprilianti@gmail.com[1]
rahmad.mahendra@cs.ui.ac.id[2]
indra@cs.ui.ac.id[3]

***Abstract*—This paper focus on building recommender system with weighted parallel hybrid method for e-commerce in Indonesia. The dataset was derived from one of the largest e-commerce company in Indonesia. The experiments used three sampling techniques, namely bootstrapping validation, timing series and systematic sampling. The best result of these experiments yields F1-measure of 9.99% which is achieved by the combination of user-based collaborative filtering approach and content-based filtering approach. Moreover, the value of evaluation metrics in this research is not much different from the previous research of recommender system. This indicates that recommender systems can be applied to e-commerce companies in Indonesia.**

***Keywords: Recommender Systems; E-commerce; Collaborative Filtering; Content-based Filtering; Weighted Parallel Hrbrid***

## I. INTRODUCTION

In recent years, the growth of electronic commerce atmosphere in Indonesia has increased rapidly since many companies have started to take a role in this business. This phenomenon can be seen in promotion and marketing offered by large e-commerce companies. The largest e-commerce companies in Indonesia including Bukalapak, Tokopedia, Elevenia, Blibli.com, Lazada, and the others. These companies are very well known among the people of Indonesia, and these companies want to increase the amount of users and transaction in their systems. Nevertheless, the companies are facing the challenges in attracting more customers to buy the products in their e-commerce website. On the other hand, the customers also have a problem in deciding which product they will buy. The later problem is often referred as the overload problem.

A system is needed to provide a reference or product information for the customers to increase their willingness to purchase something in e-commerce website. Recommender system is one of the solution to handle overload problem on the customer's side. In e-commerce companies, a recommender system can also improve the sales in three ways as follows: browsers into buyers, cross-sell, and loyalty [1].

We have seen the application of recommender system in e-commerce domain in many countries. People believe that recommender system can help e-commerce business to assist the customers in selecting products or services that are most suitable to their needs. The global e-commerce companies, such as Amazon.com and Ebay.com, had used recommender system as one of the outstanding features of their website.

The application of recommender system for e-commerce in other country may differ with the application in Indonesia. This might happen because of the differentiation of customer behavior and most popular product categories between in Indonesia and in other countries. Therefore, in this research, we implement the recommender systems for e-commerce platform in Indonesia using weighted parallel hybrid approaches, which is a combination of content-based filtering and collaborative filtering. Weighted parallel hybridization can overcome the shortcomings of those two techniques when implemented stand alone. The drawback of content-based filtering method is the lack of item variety to be recommended to the user. While, the cold start problem exists in collaborative filtering method. The system cannot provide the recommendation for new users and cannot recommend the new products to users. We also expect the performance improvement by applying hybridizations.

The rest of the paper is organized as follows: In section II, we review previous studies on developing recommender systems for e-commerce domain. Next, we explain the proposed model and implementation in section III. Section IV contains the method of experiments, the sampling techniques, and the dataset used in the experiments. Results and discussion for this paper are presented in section V, which has one figure and one table of experiment's results. Finally, in section VI we draw the conclusion and list the future works for this research.

## II. RELATED WORK

In recent years, some works have developed the recommender systems for e-commerce companies in global scope. Those studies have been conducted with different approaches, such as content-based filtering, collaborative filtering, and hybrid recommender systems [2]. For example, Amazon used collaborative filtering approach to build recommender system on their website [3]. Amazon decided to use item-based collaborative filtering instead of user-based collaborative filtering approach because it is more suitable

978-1-5090-4629-4/16/$31.00© 2016 IEEE

for large-scale retailer systems that have very large customer bases and product catalogs.

Another research that developed recommender systems using collaborative filtering approach is GroupLens Research Group from University of Minnesota [4]. They built a recommender system by using center-based neighborhood high dimension and center-based low dimension algorithm. The dataset was derived from one of e-commerce companies in US. The results achieved F1-Measure of 16.2% for center-based neighborhood high dimension algorithm and 12.8% for center-based neighborhood low dimension algorithm.

Xu et al [5], Weihong et al [6], and Kim [7] applied content-based filtering approach in developing recommender system for e-commerce domain. Kim used product taxonomy algorithm and pattern clicks for content-based filtering approach.

Hybrid approach was conducted by Pinto, Tanscheit, and Vellasco [7]. They combined item-based collaborative filtering, product positioning, and fuzzy number to build hybrid recommender system for e-commerce. The best result of this research's experiments has F1-measure more than 8%. In that research, they also showed that hybrid approach has better result than item-based collaborative filtering.

The other approaches for the recommender systems used machine learning algorithm for making recommendation, for example SVD and association rules. A recommender systems based on link prediction tested three algorithms: best-selling item, association rules, and random applied on Bizrate.com dataset [8]. Another recommender system used the item-based and SVD based on purchase data using a dataset of French retail company, à La Boite à Outils [9]. In that studies, the best-selling item algorithm can have F1-measure of 10.34%, association rules' result is 0.21%, random algorithm has result of 0.33%, and SVD's result is 10.53% for F1-measure.

Datasets of e-commerce companies tend to have high levels of sparsity and dimension of data. Therefore, the results obtained in the study of recommender system can't reach F1-measure of 20%.

To our knowledge, until this paper is done, there has been still no notable research of recommender systems applied for large scale e-commerce company in Indonesia. Meanwhile, Indonesia ranks number 5 in number of e-commerce transaction in Asia region.

## III. Proposed model

Our study aims to implement and evaluate recommender system by using weighted parallel hybrid approach which is a combination of the two approaches, collaborative filtering and content-based filtering. The proposed model consists of two phases: data modeling for collaborative filtering and content-based filtering, and implementation of the recommendations. Figure 1 shows the picture of data modeling stage.

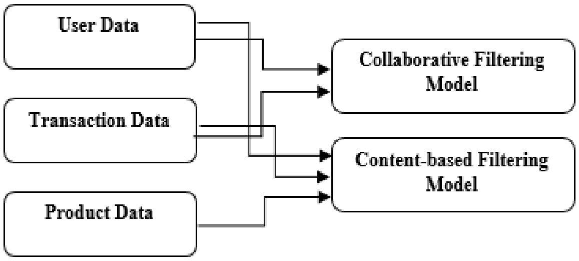

Fig. 1. The Data Modelling Stage

Figure 1 shows that the collaborative filtering model can be generated from user data and transaction data. On the other hand, content-based filtering model can be generated from user data, transaction data, and product data. After obtaining collaborative filtering (CF) and content-based filtering (CBF) data model, we can go to the implementation stage. The later stage contains the recommendation result from each approach, weighted values, and the new output from weighted parallel hybrid approach. Figure 2 shows the scheme of implementation stage.

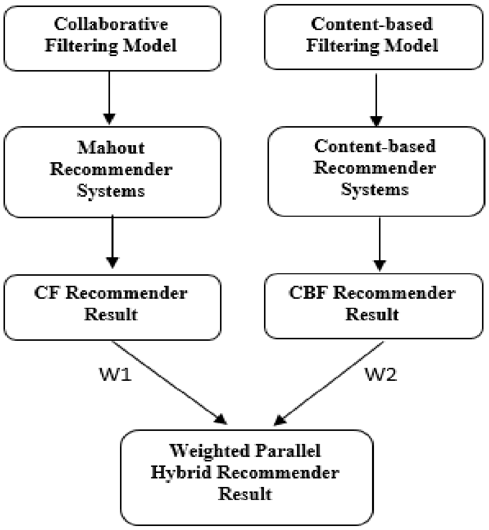
Collaborative Filtering Model
Content-based Filtering Model
Mahout Recommender Systems
Content-based Recommender Systems
CF Recommender Result
CBF Recommender Result
W1
W2
Weighted Parallel Hybrid Recommender Result

Fig. 2. The Implementation Stage

Both collaborative filtering and content based approaches have their own recommendation result. By experimenting weight values W1 and W2, the hybrid system will generate the final recommendation result.

### A. Collaborative Filtering

In our work, the collaborative filtering approach is implemented using Java programming language and Mahout framework. We experiment both user-based collaborative filtering and item-based collaborative filtering.

### 1) *User-based Collaborative Filtering*

The implementation of user-based collaborative filtering consists of three steps, namely computing similarity, determining the neighborhood, and making a recommendation.

- Computing Similarity: We compute similarity between the target user and all other users who are still active in the system. We use log-likelihood similarity algorithm in Mahout.
- Determining Neighborhood: To determine the limit of the number of users in the neighborhood, we apply user similarity by threshold value. After experiment, we choose 0.7 as the threshold neighbor value. The system will select the products that have been purchased by neighbor but not yet by the target user.
- Making a Recommendation: The products that have been selected in the previous step will be sorted based on the prediction rating given by the target user. After that, top-N products with the highest value will be recommended to the target user.

### 2) *Item-based Collaborative Filtering*

This approach is quite different with user-based collaborative filtering because the system does not seek the neighbor. Item-based collaboration filtering approach consists of two steps, namely computing similarity and making a recommendation.

- Computing Similarity: We compute similarity with log-likelihood similarity algorithm between the products that have been purchased and other products that have never been purchased by the user. The products will be sorted based on predictions rating given by user.
- Making a Recommendation: Top-N products with the highest rating value will be recommended to the target user.

### 3) *An Example of a Collaborative Filtering*

To explain more about the implementation process, we make some examples of data transaction and how each method generates recommendation from the data. The following example shows a clear difference between the implementation of user-based collaborative filtering and item-based collaborative filtering.

Suppose that we have four users and seven items in the system. We want to give an item recommendation to user A, who has not purchased Item 6 and Item 7.

Table I shows the examples of data transactions between users and the items in the systems.

TABLE I. EXAMPLES OF TRANSACTION DATA BETWEEN USERS AND ITEMS

| | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 | Item 6 | Item 7 |
|---|---|---|---|---|---|---|---|
| **A** | 0 | 1 | 0 | 1 | 1 | ? | ? |
| **B** | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| **C** | 1 | 1 | 0 | 1 | 1 | 0 | 1 |
| **D** | 0 | 1 | 0 | 0 | 1 | 1 | 0 |

In Table I, “0” value is assigned to the users who have purchased the item, “1” to the users who have never purchased the item, and “?” to the users who have never seen the item. The systems need to give the item recommendations for user A. Item 6 and item 7 have not been purchased by user A, so the recommendation will be between these two items.

- User-based Collaborative Filtering

  The log-likelihood similarity between user A and the other users in the system is shown in Table II.

TABLE II. USER SIMILARITY

| | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 | Similarity to A |
|---|---|---|---|---|---|---|
| **A** | 0 | 1 | 0 | 1 | 1 | 0.75 |
| **B** | 0 | 0 | 1 | 1 | 0 | 0.02 |
| **C** | 1 | 1 | 0 | 1 | 1 | 0.49 |
| **D** | 0 | 1 | 0 | 0 | 1 | 0.55 |

  After we get the similarity, we can find the neighborhood among the users. For this example, we choose 0.5 as threshold of neighbor. With that threshold, user D will be counted as neighbor of user A. If we looked at table I, user D bought the item 6 instead of item 7, so we will give item 6 as the recommendation product. As a result, the system will **recommend item 6** that has been purchased by user D but not by user A.

- Item-based Collaborative Filtering

  In this approach, the system just need to compute log-likelihood similarity between purchased items of user A and the other items. Table III shows item similarity between a group of purchased items (Item 2, Item 4, and Item 5) and the group of items which have not been purchased by user A (Item 6 and Item 7).

TABLE III. ITEM SIMILARITY

| | Item 6 | Item 7 |
|---|---|---|
| **Item 2** | 0.32 | 0.32 |
| **Item 4** | 0.32 | 0.63 |
| **Item 5** | 0.32 | 0.32 |
| **Total Similarity** | 0.96 | 1.27 |

Table III shows that item-based collaborative filtering **recommend Item 7** to User A. The total item-similarity generated by Item 7 is higher than Item 6.

### *B. Content-based Filtering*

In our work, the content-based filtering is implemented using Java programming language. Content-based filtering approach needs two types of data models to produce a recommendation, namely user-model and item model. User model is a model used to represent the profile of the users in the system and the item-model is used to represent the products in system. The user-model used in this research is a collection of the names of the products that have been purchased by the user, whereas the item-models only contain the name of the products. These two models are very important formed into a class in java programming to facilitate the implementation process. After that, for generating a recommendation, we need to compute similarity between user-model and item-model. We use cosine similarity formula to compute similarity in content-based filtering approach.

The cosine similarity formula is shown by Equation (1)

$$Sim(i,j) = \cos(\vec{\imath},\vec{\jmath}) = \frac{\vec{\imath}.\vec{\jmath}}{||\vec{\imath}||_2 * ||\vec{\jmath}||_2} \quad (1)$$

In Equation (1), i is the user-model and j is the item-model that are turned into vectors to compute similarity between these two models. After computing similarity between user model and item model, the system will recommend the item with the highest similarity to the user.

TABLE IV. EXAMPLE OF DATA PRODUCTS

| | Product Name |
|---|---|
| **Item 1** | Nutella & Go |
| **Item 2** | Casio Original 44DX |
| **Item 3** | Apple Ipad Mini |
| **Item 4** | Iphone 4S |
| **Item 5** | Woman V-neck |
| **Item 6** | Woman Polo Shirt |
| **Item 7** | Casio Original 4SXY |

Table IV shows the example of data products in the system. Sample data consists of seven products and each product has a unique product name.

Suppose that user A has bought Item 2, Item 4, and Item 5 in the system. Suppose also that the user A has decided not to buy Item 1 and Item 3. The system should provide recommendation to the user from the remaining products, which are Item 6 and Item 7. From that data, we can generate the user-model for user A.

Figure 3 shows the user-model for user A in the system.

$P_A$ = <"Casio Original 44DX Iphone 4S Woman V-neck">

Fig. 3. User-Model for User A

The user-model $P_A$ contains the product names that have been purchased by user A. We can compute the similarity between user-model and item-model (product name) to generate recommendation.

TABLE V. SIMILARITY WITH CBF METHOD

| | Cosine Similarity |
|---|---|
| **Sim ($P_A$ , Item 6)** | 0.2182 |
| **Sim ($P_A$ , Item 7)** | 0.4364 |

Table V shows that the cosine similarity between user-model A and Item 7 yields greater value than Item 6. As a result, the system will **recommend Item 7** for user A.

### *C. Weighted Parallel Hybrid*

Hybrid weighting is a hybridization technique that combines the recommendation resultfrom several approaches by giving weight to each approach and summing these weights to produce a new output recommendation [10]. There are two ways to determine the weight, namely empirical bootstrapping and dynamic weighting.

The formula used to generate recommendations is given in Equation. (2).

$$rec_{weighted}(u,i) = \sum_{k=1}^{n} \beta_k \; x \; rec_k \; (u,i) \quad (2)$$

In Equation. (2) $rec_{weight}$ is the recommendation results from weighted parallel hybrid approach, $rec_k(u,1)$ is the recommendation results from k approach which is given the weight $\beta_k$.

To hybridize recommendation of the collaborative filtering and content-based filtering approaches, we apply weight values consecutively w1 and w2 when the total value of w1+w2 is equal to 1.

## IV. Data and Experiments

The dataset in our experiments was derived from one of the largest C2C e-commerce company in Indonesia. This data was retrieved from 18-months period, from January 2014 to June 2015. The dataset was used in training phase and testing phase. It consists of 1,246 user data, 64,476 transaction data, and 95,468 product data.

Table VI shows the example of transaction data.

TABLE VI. Transactions Data

| Order Month | User ID | Product ID | Product Name |
|---|---|---|---|
| **201401** | A | Item 2 | Casio Original 44DX |
| **201402** | B | Item 4 | Iphone 4S |
| **201402** | D | Item 5 | Woman V-neck |
| **201403** | A | Item 5 | Woman V-neck |
| **201403** | A | Item 4 | Iphone 4S |
| **201403** | C | Item 1 | Nutella & Go |
| **201404** | B | Item 3 | Apple Ipad Mini |
| **201405** | C | Item 4 | Iphone 4S |
| **201406** | D | Item 2 | Casio Original 44DX |
| **201406** | C | Item 2 | Casio Original 44DX |

The experiment was conducted using 3 sampling techniques, namely bootstrapping validation, timing series, and systematic sampling. The purpose of this scenario is to determine the distribution of the data that is most suitable for use in a recommendation system using weighted parallel hybrid approach.

In bootstrapping validation will represent the distribution of random data, timing series will represent the distribution ordered data and the last technique, systematic sampling which is a combination of the two previous techniques will represent the distribution of random data and ordered data. The expected result in this experiment is the evaluation from the recommender system can show that the distribution of random data or ordered data cannot maximize the results of precision, recall or F1-measure.

### A. Bootstrapping Validation (BV)

Sampling is performed by separating the testing data and training data randomly from the dataset. Testing data is 30% of the overall transactions done by each user.

### B. Timing Series (TS)

This sampling technique obtains the testing data and training data based on the time of purchase (month) in period of 18 months. Testing data contain all transactions from the users who have made purchase during the last 4 months, while the training data is all transaction during the first 14 months.

### C. Systematic Sampling (SS)

Systematic sampling is dividing the testing data and training data based on specific interval. We set value 2 as the interval data. Therefore, the training data is taken by odd-index, which is 0th, 2th, 4th and so on.

## V. Results and Discussion

We tested several combination of weights. We assigned weights respectively for collaborative filtering and content-based filtering: 0.5-0.5, 0.6-0.4, 0.7-0.3, 0.8-0.2, and 0.9-0.1. For each weights, we calculated the precision, recall, and F1-measure of the output of recommendation.

The F1-measure results of these experiments is shown in Figure 4.

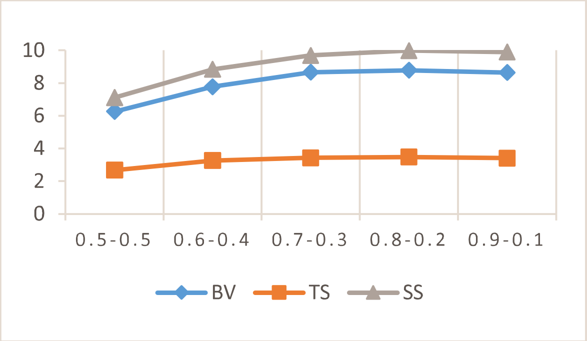

Fig. 4. F1-measure with Combination of Weights

Figure 4 shows that combination of 0.8 for collaborative filtering and 0.2 for content-based filtering generates the best result in the experiment phase. Hence, we choose 0.8 - 0.2 as weight for hybridization process.

Complete results of the experiment phase can be seen in Table VII. From table VII, we figure out that the highest average value of precision, recall and F1-measure is obtained from user-based collaborative filtering and content-based filtering with systematic sampling technique.

TABLE VII. Experimental Results

| | | Precision | Recall | F1-Measure |
|---|---|---|---|---|
| User-based CF + CBF | BV | 0.1060 | 0.0749 | 0.0878 |
| | TS | 0.0371 | 0.0327 | 0.0348 |
| | **SS** | **0.1650** | **0.0716** | **0.0999** |
| Item-based CF + CBF | BV | 0.1024 | 0.0698 | 0.0830 |
| | TS | 0.0358 | 0.0338 | 0.0348 |
| | SS | 0.1449 | 0.0602 | 0.0851 |

We can see from Table VII that not all sampling techniques can generate good results. In this research, timing series technique generates F1-measure values below 5%. It shows that the selection of sampling by order is not the right choice for e-commerce dataset.

Otherwise, the selection by random sampling can give good results from the experiment phase. However, the highest result in this study was obtained from systematic sampling techniques. Basically, systematic sampling used the combination of order sampling and random sampling, so it can get better results than the other sampling techniques.

## VI. Conclusion and Future Work

Implementation of weighted parallel hybrid approach can be applied to build recommendation system for e-commerce in Indonesia. In this research, we achieve the highest F1-measure 9.99%. This evaluation result is not much different with the result from previous research of recommendation system in global scope.

For the future work, we may explore more detail of product information, such as product categories, price, description (color, size, and quantity), to be used as attributes in content-based filtering approach. Moreover, we can try other hybridization methods such as monolithic hybrid and pipelined hybrid.

## References

[1] J.B. Schafer,J.A. Konstran, and J.T. Riedl, "Recommender System in e-Commerce", In Proceedings of ACM e-Commerce, 1999.

[2] F. Ricci, L. Rokach, and B. Shapira, "Introduction to Recommender Systems Handbook", New York: Springer, 2011.

[3] G. Linden, B. Smith, and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering", IEEE Internet Computing , vol. 7, no. 1, pp. 76–80, 2003.

[4] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Analysis of Recommendeation Algorithms for E-Commerce", In Proceedings of the 2nd ACM Conference on Electronic Commerce (EC'00). ACM, New York. pp. 285–295, 2000

[5] B. Xu, M. Zhang, Z. Pan, and H. Yang, "Content-Based Recommendation in e-Commerce", in Proceedings of Computational Science and Its Applications (ICCSA), Singapore, pp. 946 - 955, 2005.

[6] H. Weihong, andC. Yi, "An e-Commerce Recommender System Based on Content Based Filtering", Wuhan University Journal of Natural Sciences, Vol. 11, No. 5, pp. 1091-1096, 2006

[7] Y.S. Kim, "Recommender System Based on Product Taxonomy in e-Commerce Sites", Journal of Information Science and Engineering 29, pp. 63-78, 2013.

[8] M.D. Bahabadi,A.H. Golpayegani, and L. Esmaeili, "A Novel C2C E-Commerce Recommender System Based on Link Prediction: Applying Social Network Analysis", International Journal of Advanced Studies in Computer Science & Engineering, 2014.

[9] B. Pradel, S. Sean, J. Delporte, S. Rouveirol, N. Unusier,F. Fogelman-Soulie,and F. Dufau-Joel, "A Case Study in a Recommender System based on Purchase Data," Seventeenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 377–385, San Diego, 2011

[10] R. Burke, "Hybrid Recommender Systems: Survey and Experiments" USA: Journal User Modeling and User-Adapted Interaction, 2002. Journal User Modeling and User-Adapted Interaction, vol 12 issue 4, pp. 331 - 370, 2002