Computer Engineering and Applications Vol. 8, No. 2, June 2019

Book Recommender System Using Genetic Algorithm and Association Rule Mining

Hani Febri Mustika, Aina Musdholifah

Department Computer Science and Electronics, Universitas Gadjah Mada
hanifmustika@gmail.com, aina_m@ugm.ac.id

ABSTRACT

Recommender system aims to provide on something that likely most suitable and attractive for users. Many researches on the book recommender system for library have already been done. One of them used association rule mining. However, the system was not optimal in providing recommendations that appropriate to the user's preferences and achieving the goal of recommender system. In order that, this research proposed a book recommender system for the library that optimizes association rule mining using genetic algorithm to handle the previous problem. Data used in this research has taken from Yogyakarta City Library during 2015 until 2016. The experimental results of the association rule mining (Apriori Algorithm) study show that 0.01 for the greatest value of minimum support and 0.4359 for the average confidence value due to a lot of data and uneven distribution of data. Furthermore, other results are 0.499471 for the average of Laplace value, 30.7527 for the average of lift value and 1.91534252 for the average of conviction value, which those values indicate that rules have good enough level of confidence, quite interesting and dependent which indicates existing relation between antecedent and consequent. Optimization using genetic algorithm requires longer execution time, but it was able to produce book recommendations better than only using association rule mining. In addition, the system got 77.5% for achieving the goal of recommender system, namely relevance, novelty, serendipity and increasing recommendation diversity.

Keywords: Recommender System, Genetic Algorithm, Association Rule, Optimization.

1. INTRODUCTION

Recommender system development is one of the rapidly growing information technologies lately. Recommender system is a software and engineering tools which give users suggestions of items the user might want it [1]. Moreover, book recommender system includes application of recommender system. Recommendation is needed for infrequent readers. There are many methods using in recommender system. Collaboative filtering is based on the idea that a user will be interested in the same item as other users who have similar interests to the user's. A content-based recommender takes into account that a user will like items similar to the ones the userliked in the past. A knowledge-based recommender suggests items matching a user's needs, based on specific domain knowledge about how certain item features meet particular needs [2].

Several studies have used association rule mining to develop book recommender system. Association rule mining is a common technique used to identify patterns of relationships through rules. Research examined book recommender system for digital library using association rule mining to facilitate searching operation a book [3]. As a result, the system indicated that users satisfied with book recommender system. However, the system has a weakness when distribution of testing data is uneven or transactions only have one or two books, so the recommendations are hard to find and a bit. Book recommender system using association rule mining is less than optimal in providing recommendations. Consequently, it is lacking ability to show the user preferences and can't reach the goal of the recommender system. The goal of recommender system in operationally and technically is the relevance, novelty, serendipity and increasing recommendation diversity [4]. Our contribution in this paper is optimization association rules using genetic algorithm to provide recommendations and achieve the goal of the book recommender system.

2. RELATED WORK

Recommender system for library or digital library is one of interested research topics. Collaborative filtering and content-based filtering is often implied in recommender system such as book recommender system. Previous study [5] proposed collaborative filtering for digital libraries. Nevertheless, there is cold-star problem.

Other research has proposed book recommender system in school library using loan record compare three models [6]. There were SVM, Random Forest and Adaboost. The result show that the best model using SVM and the perfomance measurement using PPE (Percentage of Positive Evaluation) though quisionary.

Hybrid system is used in recommender system to optimization and handling the weakness of the previous system. Further, optimization used genetic algorithm. Genetic algorithm is optimization and search techniques. Optimization genectic algorithm based natural selection. A study about the music data recommender system combined two methods, content-based filtering and genetic algorithm [7]. The research stated that a recommender system must be able to accurately identify the trend of user preferences and provide adaptive recommendations in efficient time. The study was continued using genetic algorithm, so the new recommendation can be generated optimally every execution time, because of mutation process [8].

Based on previous studies, genetic algorithm optimization in recommender system can be combined with the other method. This study tries to optimize the use of association rule mining in book recommender system using genetic algorithm in order to provide book recommendations more optimal.

3. METHODOLOGY

3.1 DESIGN OPTIMIZATION SYSTEM

The design of book recommender system aims to the members can obtain a book which matches with their's preferences and interests. The system is designed using genetic algorithm and association rule mining.

In Figure 1, the system is divided into two parts, the first is association rule mining part and the second is genetic algorithm optimization part. Association rule

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

mining will process transaction data into rules which describe the pattern of borrowing books using Apriori algorithm. The pattern of borrowing books can be used as a reference for recommending the book to the members. Rules are stored in a database, then it is used in the process of genetic algorithm optimization.

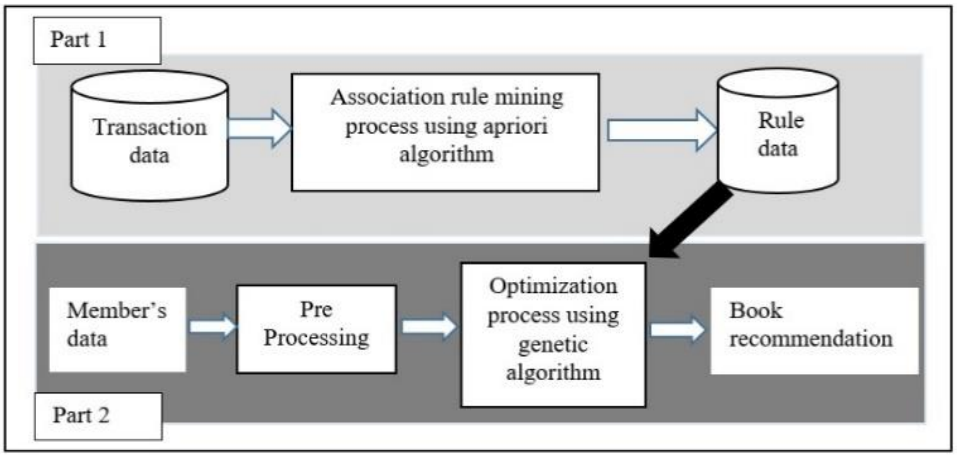

FIGURE 1. Design optimization book recommender system

Part of genetic algorithm optimization process provides optimal book recommendations for members. Member's data is input, then it is processed in pre-processing for customized data before entering the optimization process. The next process uses genetic algorithm to generate recommendations which will be borrowed book. During genetic algorithm process, evaluation process using fitness function is based on two calculations. First, the calculation uses the value of the rule that has been produced previously by the association rule mining process. Second, the calculation of the similarity is between book in the member's data which is recommended. The evaluation merging association rule mining and genetic algorithm optimization are aims to generate optimal book recommendations and to achieve the goal of the recommender system.

3.2 DESIGN PROCESS

The design processes this study using a combination of genetic algorithm and association rule in the book recommender system is shown in Figure 2. The flow diagram in Figure 2 describes process flow of book recommender system divided into two parts, there are association rule mining process and genetic algorithm optimization process**.**

In Figure 2 describes flow diagram of book recommender system. There are two parts namely association rule mining process and genetic algorithm optimization process. Association rule mining has 2 step namely frequent itemset generation and rule generation.

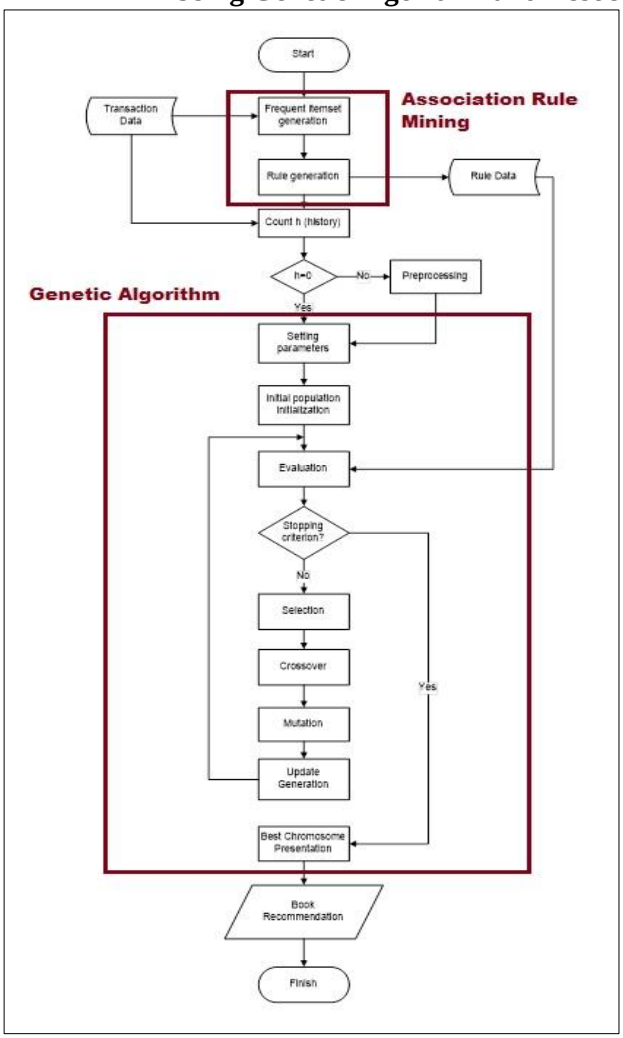

FIGURE 2. Flowchart book recommender system

Association rule mining uses the apriori algorithm to do all of the steps. Frequent itemset generation is a process which generates transaction data to frequent itemset. The frequent itemset is itemset that has support ≥ minimum support (minsup). Support can be calculated as follows

$$support\ (X \rightarrow Y) = \frac{\sigma\ (X \cup Y)}{N} \quad (1)$$

Rule generation generates rules from frequent itemset, then stored in storage. Rules which selected is rules that have confidence ≥ minimum confidence (minconf). Confidence can be calculated as follows

$$confidence\ (X \rightarrow Y) = \frac{\sigma\ (X \cup Y)}{\sigma\ (X)} \quad (2)$$

After association rule mining process, the next process is the process of calculating history (h). History (h) is the amount of book borrowing history owned by members. If history is empty (h = 0), it means that the members never borrowed book and the members have no reference to determine book recommendations, then the process is continued directly to genetic algorithm process. If history is not empty

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

(h ≠ 0), it means that the member has a history as a reference to determine book recommendations, then the history is continued pre-processing to retrieve data and information as a reference to recommend the book.

Furthermore, genetic algorithm optimization process begins with setting parameters that required by genetic algorithm. The regulated parameters include chromosome length (g), many members of the population (p), many generations or iterations (i), fitness function *f(k)*, the probability of mutation (p_m) and the probability of crossover (p_c).

Then the next process is the initialization of the initial population. Initialization of the initial population totaled p individual chromosomes and each chromosome has g genes. Encoding value of book code used to initialize genes in each chromosome. Moreover, initialization chromosome must conform a few rules, such as

- Books represented by code book, excluded books which the member ever borrow.
- Books represented by code book, excluded books which are being borrowed (unavailable).
- Books that is represented by code book, is books which appropriate a reference from history.

In the population, there is an evaluation process to calculate the fitness value of each chromosome using fitness function *f(k)* which previously arranged. The fitness function has two conditions when history (h) is not empty (h ≠ 0) and history is empty (h = 0). The fitness function can be calculated as follows

$$f(k) = \begin{cases} score\ similarity + score\ rule\ (h \rightarrow g)\ \ if\ h \neq 0 \\ score\ rule\ (contain\ g)\ if\ h = 0 \end{cases} \quad (3)$$

where h is count borrowing history, r is count rules, g is genes, x_g is book in genes g and y_j is book in history j.

When no history (h ≠ 0), the fitness value is calculated from score similarity and score rule from association rule. Similarity using cosine similarity is based on title, author's name and class code category. Using support and confidence, rule can be calculated as follow [9]

$$rule\ (x, y) = support(x, y)\ \text{x}\ confidence\ (x, y) \quad (4)$$

Stopping criterion used the number of generations or iterations (i). Genetic algorithm will stop when it fulfill maximum generations. The selection process is a process which selects chromosome in population as a parent for crossover process. Roulette wheel used as selection technique is selected chromosomes from a mating pool with a probability proportional to its fitness.

Getting parent chromosome, crossover process produces offspring chromosome or new chromosome. The next process, mutation, uses random value selection technique to fill the mutated gene. Crossover and mutation are done by referring to the probability of crossover (p_c) and the probability of mutation (p_m).

All of the offspring chromosomes from genetic operation included on the next generation to update generation. Moreover, elitism also uses in update generation process to keep chromosomes which a good fitness value so that it is not lost from the population in the next generation.

The best book recommendation is the best chromosome presentation that has the best fitness value. Each gene represents a book recommendation. All book recomendations is displayed in the next process and the process book recommendation process flow is complete. The end of the process an optimal book recommendation is produced and it achieves the goal of the recommender system.

4. RESULT AND DISCUSSION

4.1 RESULTS RULES FROM ASSOCIATION RULE MINING

Research using transaction data that has a number of books ≥ 11 per transaction. After frequent itemset generation and rule generation, the number of frequent itemset and the number of rules which was formed shown in Table 1.

In Table 1. rule was first formed when minsup 0.01 with 2 rules. The number of frequent itemset and the number of rules increased when minsup and minconf was used smaller like minsup 0.006 and minconf 0.1.

TABLE 1.
Results Frequent Itemset Generation and Rule Generation

Minimum Support (minsup)	Minimum Confidence (minconf)	Fi k=1	Fi k=2	Fi k=3	Fi k=4	Fi k=5	Rule
0,01 (Support count = 17)	0,1	173	1	0	0	0	2
	0,2	173	1	0	0	0	2
	0,3	173	1	0	0	0	2
	0,4	173	1	0	0	0	2
	0,5	173	1	0	0	0	1
	0,6	173	1	0	0	0	1
	0,7	173	1	0	0	0	0
	0,8	173	1	0	0	0	0
	0,9	173	1	0	0	0	0
0.006 (Support count = 10)	0,1	670	19	0	0	0	38
	0,2	670	19	0	0	0	37
	0,3	670	19	0	0	0	27
	0,4	670	19	0	0	0	15
	0,5	670	19	0	0	0	8
	0,6	670	19	0	0	0	2
	0,7	670	19	0	0	0	2
	0,8	670	19	0	0	0	0
	0,9	670	19	0	0	0	0

In Table 2, average confidence decreases when minsup decreases. Average confidence overall is only 0.4359. In addition to measurement of confidence, rules can be tested with others measurement methods such as Laplace, lift and conviction.

Table 3 shows results from testing measurement of rules overall. The average of Laplace is 0.499471 and the average of lift is 30.7527, which indicate that rules are quite interesting and dependent. The average value of conviction is 1.91534252, so rules are interesting. Whereas, the average of support is 0.00725 and the average of confidence is 0.4359.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

TABLE 2.
Average Confidence of Rule

Minsup	Average Confidence
0,01	0,5711
0,009	0,5711
0,008	0,533
0,007	0,4534
0,006	0,3826
0,005	0,3853
0,004	0,3726
0,003	0,3393
0,002	0,3147
Average Confidence Overall	0,4359

TABLE 3.
Results Measurement of Rule

Minsup	Average value				
	Laplace	*Lift*	*Conviction*	*Support*	*Confidence*
0,01	0,500599	27,80283	2,37374622	0,011480363	0,5711
0,009	0,500599	27,80283	2,37374622	0,011480363	0,5711
0,008	0,500412	29,35759	2,16932385	0,009969789	0,533
0,007	0,49973	27,66935	1,86486575	0,008006042	0,4534
0,006	0,498928	23,07841	1,66742302	0,006964541	0,3826
0,005	0,498842	25,64107	1,70745753	0,006181734	0,3853
0,004	0,498642	35,22133	1,79891568	0,00490051	0,3726
0,003	0,498554	36,7873	1,65735627	0,003547245	0,3393
0,002	0,498936	43,41355	1,62524812	0,002772091	0,3147
Average	0,499471	30,7527	1,91534252	0,007255853	0,4359

4.2 DIFERENCES BETWEEN RESULTS RECOMMENDATION

General differences between book recommender system using genetics algorithm and association rule mining and only using association rule as follow

- A number of books recommended by recommender system only using association rule mining are few books and hard to find because of a lot of data and uneven data distribution. While using genetic algorithm and association rule mining can more stable in providing book recommendations.
- Execution time using genetic algorithm and association rule takes longer time because the complexity of algorithm is larger.
- In achievement the goal of recommender system such as relevance, novelty, serendipity and increasing recommendation diversity, book recommender system using genetic algorithm and association rule mining is better achievement. After survey to users, book recommender system using only association rule mining gets 50.25% for achievement the goal. Book recommender system using genetic algorithm and association rule mining gets 77.5% for achievement the goal.

5. CONCLUSION

Developing and optimizing association rule mining on the book recommender system for library using genetic algorithm produce optimal book recommendations. Moreover, the recommender system using genetic algorithm and association rule mining is better achievement the goal of the recommender system such as relevance, novelty, serendipity and increasing recommendation diversity than recommender system only using association rule mining. The fitness value from book recommendations increases as long as iterating is added but execution time requires a longer time.

Based on result of research, a lot of data and uneven data distribution begin to produce rules in minsup 0.01 and the average confidence is 0.4359. Other results are 0.499471 for the average of Laplace value, 30.7527 for the average of lift value, which those values indicate that rules are quite interesting and dependent between antecedent and consequent. 1.91534252 for the average of conviction value shows that rules are enough interest.

ACKNOWLEDGEMENTS

This research was supported by Lembaga Pengelola Dana Pendidikan (LPDP) RI. We are thankful to Yogyakarta City Library (Perpustakaan Kota Yogyakarta) who has given permission to used data.

REFERENCES

[1] F.Ricci, L.Rokach and B. Shapira, *Recommender Systems Handbook*, 2nd. ed. New York: Spinger, 2015.
[2] S. Yada, "Development of a book recommendation system to inspire infrequent readers, " in *The Emergence of Digital Libraries – Research and Practices*, 2014, pp. 399-404.
[3] P. Jomsri, "Book recommendation system for digital library based on user profiles by using association rule," in *Proceedings of The Fourth Edition of The International Conference on The Innovative*, 2014, pp. 130-134.
[4] C. C. Aggarwal, *Recommender Systems*, New York : Spirnger, 2016.
[5] D.Y. Patil, A.Shekh and G. Kadlag, "Survey on Collaborative Filtering for Digital Libraries," *International Research Jorunal of Enggineering and Technology*, vol. 3, issue 12, pp. 1013-1016, December 2016.
[6] K. Tsuji, F. Yoshikane, S. Sato, and H. Itsumura, "Book recommendation using Machine learning methods based on library loan records and bibliographic information," *International Journal of Academic Library and Information Science*, vol. 3, pp. 7-23, January 2015.
[7] G.P. Wakure and V. Kadrolli, "Music recommendation system using genetic algorithm," *International Journal of Engineering Research and Technology (IJERT)*, vol. 2, pp. 580-582, September 2013.
[8] N. Badhe, D. Mishra, C. Joshi, and N. Shukla, "Recommender system for music data using genetic algorithm," *International Journal of Innovations and Advancement in Computer Science*, vol. 3, pp. 66-69, April 2014.
[9] D. Jannach M. Zanker, A. Felfernig, and G. Friedrich, *An Introduction Recommender Systems*, New York: Cambridge University Press, 2011.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)