KICS The Korean Institute of Communications and Information Sciences

Available online at www.sciencedirect.com

ScienceDirect

ICT Express 6 (2020) 214–219

www.elsevier.com/locate/icte

Adaptive rule: A novel framework for recommender system

Erna Hikmawati*, Nur Ulfa Maulidevi, Kridanto Surendro

School of Electrical Engineering and Informatics, Bandung Institute of Technology, Bandung, Indonesia

Received 28 February 2020; received in revised form 20 May 2020; accepted 5 June 2020
Available online 13 June 2020

Abstract

Determining minimum support value on association rule is not easy to users. Hurdles such as data locations and data origins are a mountain that should be overcome by association rule. In addition, rules generated do not necessarily fit user's needs, thereby, requiring the rules to be put in order based on certain criteria. This paper will discuss a model to produce adaptive rules that originate from a number of different databases, automated threshold determination and rule ranking based on predetermined criteria. The adaptive rule model produced can be implemented in recommender systems in various fields.

Keywords: Adaptive rule; Association rule; Recommender system

Contents

1. Introduction

Current technological advancement has caused a massive increase in data growth. While the data may contain useful information, it has to be extracted using certain methods. One of known methods to obtain information is data mining technique. With data mining, valuable knowledge or information from a cluster of big data can be retrieved [1–4]. Consequently, the information gained may be utilized in decision making process.

One of the functions of data mining is Association Rule. Association Rule, or also known as Frequent Itemset Mining, is a basic concept in data mining and the most common method to identify association relationship based on frequent occurrence of itemset [1,2,5,6].

In Association Rule, the first thing to do is to determine a threshold called minimum support. Minimum support functions to determine which item will be included in rule establishment process and which item are not observed from the frequency of item occurrence.

* Correspondence to: School of Electrical Engineering and Informatics, Ged. Achmad Bakrie,Lt. 2, Jl. Ganesha No 10, Bandung 40132, Indonesia.
E-mail addresses: erna@pasim.ac.id (E. Hikmawati), ulfa@informatika.org (N.U. Maulidevi), endro@informatika.org (K. Surendro).

Peer review under responsibility of The Korean Institute of Communications and Information Sciences (KICS).

https://doi.org/10.1016/j.icte.2020.06.001

Determining minimum support represents an intricate task for users, because minimum support value significantly affects the rules to be established. A minimum support value that is too high causes the items involved in the rule formation to be too small, leading to high probability of missing information. Conversely, a minimum support value that is too low causes the items involved in the rule formation to be too big, consuming too much time and memory [4,6].

The next hurdle is heterogeneity. Existing data size and complexity require innovative solution from the field of Information and Communication Technology [7]. Data mining should be able to handle heterogeneous data [7]. The growth of various data from different sources requires rule establishment to be able to integrate its various source databases [8]. There are several studies conducted attempting to solve the issue, one of which is using Meta Association Rule method. Meta association rule is a rule about rules build upon regular association rules that have been previously extracted from a set of databases with similar topic and structure [9]. However, meta association rule method is unable to establish rules from databases with different characteristics and attributes. Therefore, there is a glaring need of a new method that can generate rules from databases with different sources, characteristics or attributes.

In order to come to terms with current condition that is rife with volatility, uncertainty, complexity and ambiguity or also known as the VUCA world, decision makers are demanded to make appropriate strategic choice [10]. Current challenge is how to overcome and respond creatively to the challenges of the VUCA world [11]. Present-day disruption era is changes that render previous products, services, and processes ineffective [12]. Adaptive strategies must be considered in taking on the future [13]. Decision makers have different criteria in determining what information they need, thereby, rule establishment must consider criteria desired by its users which ultimately results in Adaptive Rules [14]. Adaptive Rules generated can be utilized in recommender systems.

Nowadays, recommender systems are widely implemented in various sectors and have generated significant benefits [15]. For example, in e-commerce, the system will display the next most bought item at a time to entice customers to buy it too. Another example, Google has also implemented recommender systems in its search engine, where Google will show recommended words most searched by people. Recommender systems can also be implemented in medical field to detect the probability of a disease to occur based on patients' medical records. Furthermore, based on the record of drugs administered to a patient, recommender systems can display proper drugs to consume by combining through the contradictions of every available drug.

Collaborative Filtering and Content-Based Recommendation methods are the most common methods used in recommender systems [15–18]. In Collaborative Filtering (CF) method, user profile, user preference and user behavior are of significance importance. There are two analysis subjects in CF namely explicit and implicit characteristics. Explicit characteristics include ratings on items, items viewed, items purchased and so on. While implicit characteristics include how long users interact with the content and the level of the interaction [15]. However, for this method, issues arise when new users pop up devoid of any behavior, or when new items with no ratings are registered, and when a system is programmed with high level of privacy [15]. One solution is to use Association Rule Mining on Recommender Systems. Several studies mention that Association Rule Mining performs better compared to CF for Recommender Systems [15–18].

This paper will discuss models to generate adaptive rules that originate from different multiple databases, both external and internal, automated threshold and involve various criteria to assess rule priority. Adaptive rule model produced will be implemented in recommender systems in various fields and sectors. This paper is structured as follows: Related Studies will be presented in Section 2. Discussion will be presented in Section 3, which is methodology suggestions in this journal.

2. Related work

This section will look closer at methods of minimum support value determination, meta association rule, multi-criteria, decision analysis and recommender systems in previous studies.

2.1. Association rule basic concept

Association Rule Mining can be defined formally as follows: if I = {i1, i2, …, im} is a set of items. D is a set of T transactions, where one set of T transactions is a set of items, then T ⊆ I. If A is a set of items. T transactions is said to contain A if and only if A ⊆ T. Association rules are A → B forms, where A ⊆ I, B ⊆ I, and A ∩ B = Ø. A → B rule has support in the set of D transaction if s% from transactions in D contains A ∪ B [19,20], where support is a ratio between transaction frequency containing itemset A divided by transaction total. In general, it can be formulated as follows [9]:

$$Supp(A \rightarrow B) = \frac{|t \in D | A \cup B \subseteq t|}{|D|} \quad (1)$$

where:

- Supp is support value
- A is an antecedent rule in the form of itemset
- B is a consequence in the form of itemset
- t is a transaction that contains A and B
- D is total transaction

A → B rule applies in the set of D transaction with confidence c if c% of the transaction in D that contains A also contains B. Confidence is a ratio between the number of transaction that contains item A and B divided by the transaction that contains A. Confidence value of A → B rule is obtained with the following formula [9]:

$$Conf(A \rightarrow B) = \frac{|t \in D | A \cup B \subseteq t|}{|t \in D | A \subseteq t|} \quad (2)$$

where:

- Conf is Confidence
- A is an antecedent rule in the form of itemset
- B is a consequence in the form of itemset
- t is a transaction that contains A and B
- D is total transaction

The general framework of association rule is extracting rules whose support and confidence values exceed the minsupp and minconf values. Minsupp and minconf values are variables determined by users. In this case, it can be said that $A \rightarrow B$ is considered in frequent category if $Supp(A \rightarrow B) \geq$ minsupp and is considered in confidence category (strong rule) if $Conf(A \rightarrow B) \geq$ minconf.

2.2. Minimum support value determination method

Association Rule Mining is first put forward by Agrawal [21]. There two initial methods of association rules namely Apriori [22] and FP-Growth [23]. In terms of minimum support determination, both use minimum support value determined by users.

Most association rules establish one threshold of minimum support for all items, however, in reality, different items may have different criteria to assess their importance. Support should vary for different items [24]. Thus, a myriad of studies shows up that propose methods for association rules that utilize different minimum support for each item called multiple minimum support [3,25]. However, implementing multiple minimum support requires users to determine minimum support for each item, adding more task to an already hard one.

Another view on minimum support value determination in association rules is to remove minimum support from the rule formation process. There are two association rule methods that do not take into account minimum support value namely Top-k and skyline. In top-k association rule method, instead of minimum support value, k value is needed to signify the number of rules to be generated [2]. With this, it is easier for users to determine k value because they explicitly know how many rule results they want. Top k association rule method is first put forward by Fournier and Cheng Wei Wu [26,27].

Skyline algorithm is first put forward by Borzsony [28] and then developed further by Goyal [29]. Skyline is defined as points that are not dominated by other points [28]. Skyline algorithm is then combined to produce association rule by Jerry Chun-Wei Lin [5] and Jeng-Shyang Pan [4]. In this study, we do not utilize minimum support but instead use utility maximal (utilmax) obtained from every iteration performed by utility list structure.

2.3. Meta association rule

A common requirement that often emerges during Association Rule Mining is obtaining rules from different sources. It has been studied by many researchers. One of them is Ruiz [9] that offers new framework to obtain association rules that cover many databases. This framework is called Meta Association Rule. Meta Association Rules are rules build upon regular association rules that have been extracted previously from a set of databases with similar topic and structure [9]. However, meta association rules can only be formed from databases that possess identical attributes.

2.4. Multi-criteria decision analysis

To determine which rule is the best, a certain method combined with association rules is needed. One suitable method for prioritizing rules is decision analysis technique [30]. This approach can help decision makers in determining which rules is the most suitable and is ranked higher than other rules.

2.5. Recommender systems

Recommender Systems have been widely employed in various domains. While the majority of recommender systems employ collaborative filtering, a new development of recommender systems that uses association rule mining [15–18,31]. The first recommender system based on association rule mining is conducted by Kim [17]. The study shows that the method proposed performs better than collaborative filtering.

3. Discussion and conclusion

The development of studies on association rules has led to varied minimum support determination methods. Yet, up to this point, a method that is able to determine minimum support value in an automated manner based on its database characteristics has not been developed yet.

The rapid growth of data in terms of size causes various databases to emerge. Traditional method of association rule is becoming more and more irrelevant when it comes to mining rules that originate from multiple databases. Previous researchers have proposed a method called Meta Association Rule that is able to combine several databases in rule mining process. However, Meta Association Rule method is not yet able to combine databases with different characteristics.

Rules generated during the association rule mining process have not indicated priority or ranking of every rule, thus, all rules are considered identical. However, the ranking of rules is paramount in decision making process to identify which rule is the most relevant and which rule is based on any certain criteria.

To sum up the problems and challenges related to association rules, hereby the researcher presents the review result on a number of studies on association rules and recommender system in Table 1.

Table 1
Review result on association rule and recommender system studies.

No	Current method	Weakness	Requirement
1	Determining minimum support		
	Apriori [21,22]	Minimum support is determined by users	A method that can calculate threshold in accordance with certain database characteristics and criteria to determine which item will be included in the rule establishment process.
	Fp-Growth [23]		
	Top-k [2,26,27]	The possibility of missing information due to limited number of rules	
	Skyline [4,5,28,29]	Using umax that changes at every iteration	
	Multiple Minimum Support [3,24,25]	Users face difficulties to determine minimum support value for each item	
2	Meta Association Rule [8,9][32]	It can only extract rules from databases with similar attributes	Generating rules from several databases with different characteristics and sources
3	Multicriteria Decision Analysis [30][33,34]	Separate process from rule formation	Rule ranking process in accordance with certain criteria
4	Recommender system		
	Collaborative Filtering and Content Based Filtering [15–18]	Highly dependent on user profiles, user preferences and user behaviors	Not dependent on a certain database

To address the weaknesses and requirements laid out in Table 1, the researcher proposes a new method to generate adaptive rules like in Fig. 1.

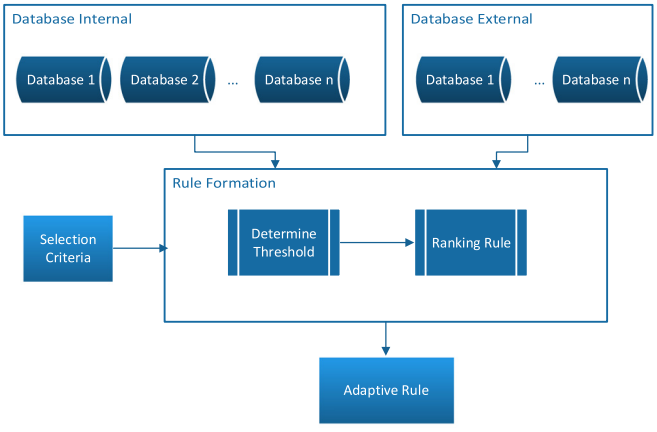

Fig. 1. Proposed methodology for adaptive rules.

This methodology is expected to produce adaptive rules with better methods than the previous method. A detailed explanation of this method is as follows:

1. Input
There are three types of input to this method, namely internal database, external database, and selection criteria. One that is different from this method can have two types of databases that are inputted namely an internal database and an external database. In the previous methods used as input is only an internal database, namely the transactional database that is directly related to the process of establishing rules. Examples of internal databases are user profiles, item descriptions, and purchase transaction history. But this method is proposed to involve an external database which in fact can be a factor for decision making. Examples of external databases are product reviews, current trends, and so on. Another input to this method is the selection criteria chosen to determine the ranking rule and item selection. In the basic association rule concept, the item selection process is only based on the frequency of occurrence of the item in a transaction. But in reality, many other factors and criteria are needed in providing recommendations. These criteria can be in the form of price of goods, profit of goods, level of importance of goods,

and so forth. In this method decision makers can determine which criteria will be used as a benchmark in the preparation of rules and ranking, so that everyone who has different criteria will get a different rule later even though it comes from the same transaction database.

2. Process
   The main process in this method is the rule formation using the basic method of association rule, but improvements have been made at several points, namely in the process of determining the minimum support value and ranking rules. The minimum support value in this method is generalized as a threshold. This is done because the selection of items in the formation of rules does not only come from the support or frequency of the item but based on other criteria determined by the user in the previous stage. After the threshold is determined for the rule formation process, the next process is the formation of a rule by considering the criteria that have been determined and ordered according to the threshold before so that the rule formed will not need to be ranked again.
3. Output
   The output of this method is an adaptive rule, a rule that adjusts to some circumstances, namely the input database and user-specified criteria. The hope is that this adaptive rule can provide appropriate recommendations if applied to the recommender system.

The proposed methodology to generate adaptive rules possess following advantages:

1. The mining of association rules that originate from different sources, both based on characteristics or topic. Recommender cannot depend on one database since there are probably many databases that can be combined to obtain rules, whether they are internal or external databases. Internal databases refer to transaction databases, user profiles, and so on. While external databases refer to data that is not directly related to the system such as current issues, review database of an item, and so on.
2. In the establishment process, adaptive rules are not only based on item occurrence frequency in transactions but also involving certain criteria chosen by users according to their interests.
3. In the proposed methodology, users are not required to determine minimum support value, minimum confidence value, and so on since it already has a dedicated function to determine the threshold of the determination of items to be involved in the rule establishment process according to chosen criteria and it takes into account the characteristics of databases inputted.
4. There is a process that prioritizes and ranks rules generated. In this process, the rules generated in the previous process will be ranked according to certain criteria determined by users. The ranking process is performed by comparing the significance of every existing rule.

The proposed method not only combines several previous methods, but also makes improvements to overcome the weakness of the previous method. Improvements made include:

1. In the proposed method will be established to determine the minimum support value. This is done because most of the research conducted focuses on improving the performance of association rules as seen from its runtime and memory needs. There is no research focused on producing models that can automatically determine the minimum value of support based on the characteristics of the dataset.
2. Overcoming the weakness of the meta association rule, namely combining databases with different characteristics and attributes. Meta Association rules can extract rules from several databases, but rules are extracted only on databases that have the same characteristics and attributes. The need now is to collaborate on different databases to get relevant and interesting information.
3. Insert the rule ranking process at the time of the rule formation so that it does not require a long time to process it. At present some researchers combine association rules with the multi-criteria decision analysis method to rank rules so that rules can be sorted from the most interesting and relevant based on certain criteria. But in previous studies this process is a separate process from the rule formation process, so it requires a lot of time to process it.
4. The process of forming this rule will later be implemented in the recommender system. There are several studies that utilize association rules in the recommender system, but the algorithm used is the Apriori algorithm which is the basic concept of the Association rule and still has many shortcomings.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Bandung Institute of Technology and Pasim National University for supporting this research.

References

[1] J.M. Luna, P. Fournier-Viger, S. Ventura, Frequent itemset mining: A 25 years review, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. (2019) http://dx.doi.org/10.1002/widm.1329.

[2] H. Ryang, U. Yun, Top-k high utility pattern mining with effective threshold raising strategies, Knowl.-Based Syst. 76 (2015) 109–126, http://dx.doi.org/10.1016/j.knosys.2014.12.010.

[3] W. Gan, J.C.-W. Lin, P. Fournier-Viger, H.-C. Chao, J. Zhan, Mining of frequent patterns with multiple minimum supports, Eng. Appl. Artif. Intell. 60 (2017) 83–96, http://dx.doi.org/10.1016/j.engappai.2017.01.009.

[4] J.-S. Pan, J.C.-W. Lin, L. Yang, P. Fournier-Viger, T.-P. Hong, Efficiently mining of skyline frequent-utility patterns, Intell. Data Anal. 21 (2017) 1407–1423, http://dx.doi.org/10.3233/IDA-163180.

[5] J.C.-W. Lin, L. Yang, P. Fournier-Viger, T.-P. Hong, Mining of skyline patterns by considering both frequent and utility constraints, Eng. Appl. Artif. Intell. 77 (2019) 229–238, http://dx.doi.org/10.1016/j.engappai.2018.10.010.

[6] Q.-H. Duong, B. Liao, P. Fournier-Viger, T.-L. Dam, An efficient algorithm for mining the top- k high utility itemsets, using novel threshold raising and pruning strategies, Knowl.-Based Syst. 104 (2016) 106–122, http://dx.doi.org/10.1016/j.knosys.2016.04.016.

[7] K. Gibert, J.S. Horsburgh, I.N. Athanasiadis, G. Holmes, Environmental data science environmental modelling & software, Environ. Model. Softw. 106 (2018) 4–12, http://dx.doi.org/10.1016/j.envsoft.2018.04.005.

[8] M.D. Ruiz, J. Gómez-Romero, M. Molina-Solana, M. Ros, M.J. Martin-Bautista, Information fusion from multiple databases using meta-association rules, Internat. J. Approx. Reason. 80 (2017) 185–198, http://dx.doi.org/10.1016/j.ijar.2016.09.006.

[9] M.D. Ruiz, J. Gómez-Romero, M. Molina-Solana, J.R. Campaña, M.J. Martin-Bautista, Meta-association rules for mining interesting associations in multiple datasets, Appl. Soft Comput. 49 (2016) 212–223, http://dx.doi.org/10.1016/j.asoc.2016.08.014.

[10] F. Giones, A. Brem, A. Berger, Strategic decisions in turbulent times: Lessons from the energy industry, Bus. Horiz. 62 (2019) 215–225, http://dx.doi.org/10.1016/j.bushor.2018.11.003.

[11] E. King, R. Badham, Leadership in uncertainty, Organ. Dyn. (2018) http://dx.doi.org/10.1016/j.orgdyn.2018.08.005, S0090261618301001.

[12] C. Millar, M. Lockett, T. Ladd, Disruption: Technology, innovation and society, Technol. Forecast. Soc. Change 129 (2018) 254–260, http://dx.doi.org/10.1016/j.techfore.2017.10.020.

[13] H.R. Maier, J.H.A. Guillaume, H. van Delden, G.A. Riddell, M. Haasnoot, J.H. Kwakkel, An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together? Environ. Model. Softw. 81 (2016) 154–164, http://dx.doi.org/10.1016/j.envsoft.2016.03.014.

[14] E. Hikmawati, K. Surendro, How to determine minimum support in association rule, in: Proceedings of the 2020 9th International Conference on Software and Computer Applications, 2020, pp. 6–10, http://dx.doi.org/10.1145/3384544.3384563.

[15] T. Osadchiy, I. Poliakov, P. Olivier, M. Rowland, E. Foster, Recommender system based on pairwise association rules, Expert Syst. Appl. 115 (2019) 535–542, http://dx.doi.org/10.1016/j.eswa.2018.07.077.

[16] P. Paranjape-Voditel, U. Deshpande, A stock market portfolio recommender system based on association rule mining, Appl. Soft Comput. 13 (2013) 1055–1063, http://dx.doi.org/10.1016/j.asoc.2012.09.012.

[17] Y.S. Kim, B.-J. Yum, Recommender system based on click stream data using association rule mining, Expert Syst. Appl. 38 (2011) 13320–13327, http://dx.doi.org/10.1016/j.eswa.2011.04.154.

[18] J. Jooa, S. Bangb, G. Parka, Implementation of a recommendation system using association rules and collaborative filtering, Proc. Comput. Sci. 91 (2016) 944–952, http://dx.doi.org/10.1016/j.procs.2016.07.115.

[19] M. Kantardzic, DATA MINING Concepts, Models, Methods, and Algorithms, second ed., John Wiley & Sons, Inc., Hoboken, New Jersey, 2011.

[20] C. Zhang, S. Zhang, Association Rule Mining: Models and Algorithms, Springer, Berlin, New York, 2002.

[21] R. Agrawal, T. Imielinski, A. Swami, H. Road, S. Jose, Mining association rules between sets of items in large databases, in: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, USA, 1993, pp. 207–261.

[22] R. Agrawal, Fast algorithms for mining association rules, in: Proceedings of the 20th International Conference on Very Large Data Bases, San Francisco, CA, USA, 1994, pp. 487–499.

[23] J.H.J. Pei, Mining frequent patterns without candidate generation, in: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, 2000, pp. 1–12, http://dx.doi.org/10.1145/342009.335372.

[24] Y.-C. Lee, T.-P. Hong, W.-Y. Lin, Mining association rules with multiple minimum supports using maximum constraints, Int. J. Approx. Reason. 40 (2005) 44–54, http://dx.doi.org/10.1016/j.ijar.2004.11.006.

[25] Y.-H. Hu, Y.-L. Chen, Mining association rules with multiple minimum supports: a new mining algorithm and a support tuning mechanism, Decis. Support Syst. 42 (2006) 1–24, http://dx.doi.org/10.1016/j.dss.2004.09.007.

[26] P. Fournier-Viger, C.-W. Wu, V.S. Tseng, Mining top-K association rules, in: L. Kosseim, D. Inkpen (Eds.), Advances in Artificial Intelligence, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 61–73, http://dx.doi.org/10.1007/978-3-642-30353-1_6.

[27] C.W. Wu, B.-E. Shie, V.S. Tseng, P.S. Yu, Mining top-K high utility itemsets, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12, ACM Press, Beijing, China, 2012, p. 78, http://dx.doi.org/10.1145/2339530.2339546.

[28] S. Borzsony, D. Kossmann, K. Stocker, The skyline operator, in: Proceedings 17th International Conference on Data Engineering, IEEE Comput. Soc., Heidelberg, Germany, 2001, pp. 421–430, http://dx.doi.org/10.1109/ICDE.2001.914855.

[29] V. Goyal, A. Sureka, D. Patel, Efficient skyline itemsets mining, in: Proceedings of the Eighth International C* Conference on Computer Science & Software Engineering, C3S2E '15, ACM Press, Yokohama, Japan, 2008, pp. 119–124, http://dx.doi.org/10.1145/2790798.2790816.

[30] D.H. Choi, B.S. Ahn, S.H. Kim, Prioritization of association rules in data mining: Multiple criteria decision approach, Expert Syst. Appl. 29 (2005) 867–878, http://dx.doi.org/10.1016/j.eswa.2005.06.006.

[31] J. Gera, H. Kaur, A novel framework to improve the performance of crowdfunding platforms, ICT Express 4 (2018) 55–62, http://dx.doi.org/10.1016/j.icte.2018.04.011.