Iyer, S. et al. (Eds.) (2022). Proceedings of the 30th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Association Analysis on Open-Ended Concept Maps using Data Mining

Didik Dwi PRASETYA[a], Setiadi Cahyono PUTRO[a], Muhammad ASHAR[a], Saida ULFA[b], & Tsukasa HIRASHIMA[c]
[a]*Department of Electrical Engineering, Universitas Negeri Malang, Indonesia*
[b]*Department of Educational Technology, Universitas Negeri Malang, Indonesia*
[c]*Department of Information Engineering, Hiroshima University, Japan*
*didikdwi@um.ac.id

Abstract: An open-ended concept map is an appealing approach that allows learners to add concepts and links freely, representing their understanding. This technique reveals differences among students and accurately captures their knowledge structure. However, manually evaluating an open-ended map is challenging and time-consuming, particularly in a large classroom, and involves many propositions. Several works tried to use data mining techniques to generate concept maps from the source text rather than analyzing the knowledge structure. The previous study employed association analysis on provided concept maps. This study focused on association analysis of open-ended concept mapping in revealing hidden students' understanding. A total of 20 open-ended concept maps were involved in this study. The findings suggested that data mining techniques could quickly disclose necessary information on open-ended concept maps. The study identified an exciting topic for further research in educational data mining, particularly concerning knowledge structure.

Keywords: Concept map, open-ended, association analysis, data mining, educational data mining

1. Introduction

Concept maps are beneficial graphical knowledge representation tools to acquire an overview of an individual's knowledge domain. The concept map is well-known for accurately representing individual conceptual knowledge in a given topic. It depicts different ideas at nodes with linking words connecting two concepts describing their relationship and forming a proposition (Cañas and Novak, 2014). Propositions are the minor semantic units and are fundamental elements of a map. Therefore, it can be perceived as a crucial knowledge component (Aleven et al., 2016). The proposition represents a unit's declarative knowledge in forming meaningful statements.

Concept map construction can be divided into open-ended and closed-ended maps (Hirashima, 2019; Prasetya et al., 2019). In an open-ended approach, learners can add concepts and link terms in their diagrams to represent their understanding. The knowledge structure in the open-ended concept map can be analogized as unstructured or indeterministic information. On the other hand, a closed-ended concept mapping style has finite components. Thus, learners must use the defined concepts and links in building their concept maps. The closed-ended concept map allows automatic assessment, provides quick feedback (Hirashima, 2019; Pailai et al., 2017), and motivates learners to reach maximum learning comprehension (Prasetya et al., 2021).

The open-ended approach is more challenging in evaluation than the closed-ended method. So far, the technique used to evaluate students' open-ended concept maps might be using rubrics and assessed by an expert (Taricani and Clariana, 2006). However, rubrics have limitations, especially in the case of various open-ended concept maps. The rubric is time-consuming and cannot comprehensively evaluate students' knowledge structure (Yoo and Cho, 2012). The diversity of concept maps makes it difficult for teachers to determine certain students' mental model patterns. However, judging their knowledge structure is an excellent choice. One approach that can be proposed to manage concept maps automatically is data mining techniques.

Data mining automatically extracts new and vital knowledge hidden in large datasets (Yoo and Cho, 2012). Data mining aims to uncover the databases' latent, previously undisclosed trends and patterns. Furthermore, the data mining process generates valuable information needed to support decision-making. Data mining has been proven to provide many benefits and is widely used in various fields, one of which is education and is called Educational Data Mining (EDM). EDM is a novel multidisciplinary research field focusing on developing methods for analyzing data generated in educational settings (Romero and Ventura, 2020). EDM is concerned with developing tools for exploring the unique types of data found in educational settings and, by doing so, better understanding students and the environments in which they learn (Baker, 2019).

Until recently, many studies have analyzed the impact of data mining in managing concept maps. For example, Villalon and Calvo (2008) revealed a task of automatically generating concept maps from a text known as Concept Map Mining (CMM). CMM is an automatic or semi-automatic creation of concept maps from text documents. The same concept was also used for the automatic generation of concept maps based on text analysis and association rule mining (Shao et al., 2020). For a similar purpose, a semi-automatic method for constructing a concept map of reading material in English for Foreign Language (EFL) was proposed and yielded positive effects (Pinandito et al., 2021). However, CMM aims to produce a concept map that provides accurate information about students' understanding instead of analyzing the concept map expressed by the students themselves.

The present study utilized data mining techniques to process open-ended concept maps and analyze students' understanding of learning topics. Another study similar to this study was conducted by Yoo and Cho (2012). However, the previous approach was intended for pre-defined concept maps rather than open-ended ones that refer to students' understanding. This study used association rules to reveal valuable hidden information in concept maps. In particular, the following research question guides the study: the extent to which data mining techniques on the concept map could quickly represent students' understanding of a Database topic.

2. Methods

2.1 Research Design

The term mining concept maps in this study refer to using data mining techniques on open-ended concept maps to identify students' understanding quickly. In particular, the association analysis methodology was applied to the concept maps data to obtain valuable hidden information. Since the interpretation of learners' open-ended concept maps cannot be made as easily as the closed method, this technique will make a non-trivial contribution to the learning process.

The association rule is a proper data mining technique for determining the extent of correlation between variables in a database (Ait-Mlouk et al., 2017). It aims to discover regularity in the data by identifying specific patterns. Association rules' main activity is finding essential relationships between items in each transaction. The relationship could reveal the extent of the strength of a rule in the association.

2.2 Data Collection

This study involved 20 open-ended concept maps made by 20 second-year university students majoring in Informatics Engineering. The concept maps were created using the scratch-build concept mapping tool by referring to the "Basisdata Relasional" (Database Relational) topic in the "Basisdata 1" (Database 1) course that was delivered in Indonesia.

The concept map extraction produced 314 concepts with an average of 15.70 (SD = 4.34) that describe the original learners' ideas regarding the Basisdata Relasional topic. Concepts represent key ideas or opinions expressed by learners related to the learning materials. The learners' concepts produced 290 propositions with an average of 14.50 (SD = 4.37).

2.3 Preprocessing

As in general data mining activities, preprocessing was carried out before conducting the main activity. Data preprocessing is an essential task and critical step in data mining (Angiani et al., 2016). It includes the preparation and transformation of data into an appropriate form. The primary purpose of preprocessing is to normalize data, reduce data size, discover the relations between data, remove outliers, and extract features for data (Alasadi and Baya, 2017). Preprocessing was applied to the pairs of concepts that form the propositions on the concept maps. It transforms the raw ideas and propositions data into a valuable and efficient format. This stage performed the initial processing of the set of propositions on students' concept maps. Initial processing includes reading and parsing concept maps from the database, data cleaning, tokenization, and term weighting.

2.4 Mining Open-Ended Concept Maps

The open-ended mining concept maps process students' knowledge structures to reveal their understanding of related learning materials. The main steps of the mining activity are shown in Figure 1. Each concept map represents a student's knowledge structure. They were treated as a single document and extracted for subsequent operations. There were two main recognition operations: concept recognition and relationship recognition. Based on the recognition operations that have been carried out, association analysis could be performed to reveal common concepts and propositions used by students on their concept maps.

Concepts recognition is the process of finding potential candidates for a student's key ideas. Each concept defined by learners is identified and retrieved from the open-ended database system. Frequent one itemset was applied to determine the emergence of ideas in each concept map. Since the concept is a unique element in the individual concept map, analysis was conducted in the group's structure. Furthermore, the concept recognition operation results could be confirmed to identify what students understand through the knowledge structure that has been formed. The support value for a one itemset X is obtained by dividing the number of transactions containing X by the total number of transactions and is defined as support (X) = PX / P total.

Relationships are propositions that describe the relationship between two concepts that are connected by a link label. Proposition represents the smallest meaningful component on concept maps. It has an important role in revealing students' understanding related to the knowledge they have acquired. In the relationships recognition phase, each concept that is fully connected to another concept is identified as a sub-concept map. The relationships are represented in the form of association rules or sets of frequent items. For example, a frequent itemset, {database, table} indicates that there was a strong relationship between database and table terms in the database material.

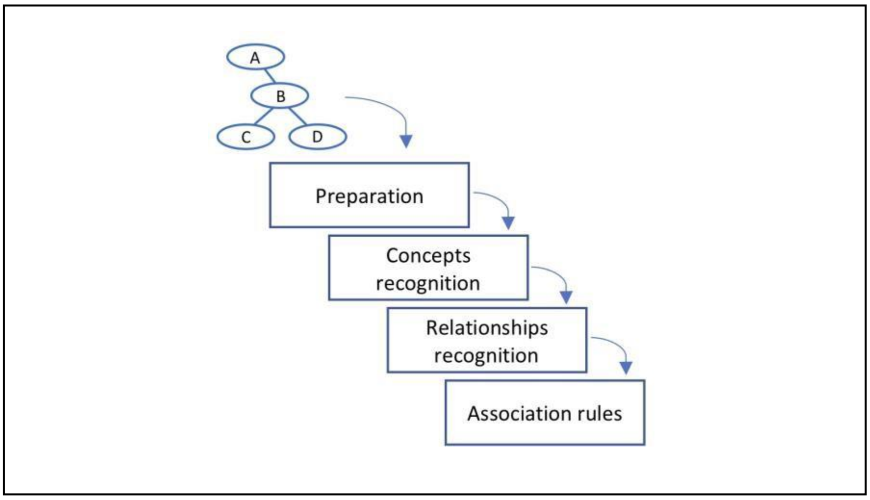

Figure 1. Mining concept map activities

Association rules uncover intriguing connections or relationships among a massive number of data items (Karabatak et al., 2011). It displays attributes value circumstances that frequently occur together in a given dataset. The main purpose of association rules is to recognize the patterns of co-occurrence of attributes in a database. The quality of association analysis results is often measured using support and confidence parameters (Xu, 2016). After receiving a frequent high value, furthermore, the confidence value was examined using associative rules. The present study employed the Apriori algorithm to analyze frequent association concept map mining. A minimum frequency threshold of 30% was used to analyze concepts, propositions, and association rules.

3. Results and Discussion

Twenty learners' concept maps were extracted from the database and generated 314 unique concepts expressing various ideas. Analysis of pattern creation was carried out on the collection of concepts by using a frequent one itemset. The purpose of this analysis was to identify the ideas that learners most express. Minimum support of 30% was set to reveal frequent concept patterns. The concept "basis data relasional/*relational database*" was the idea that most of the learners sparked. A total of 18 learners (90%) wrote it on their concept maps. Learners' concepts with a minimum support of 50% were: "atribut/*attribute*" (85%), "relasi/*relation*" (85%), "tupel/*tuple*" (85%), "domain/*domain*" (80%), "kardinalitas/*cardinality*" (70%), "tabel dua dimensi/*2 dimension table*" (65%), "derajat/*degree*" (65%), and "istilah/*terms*" (50%). Other frequent itemset of concepts were: "super key/*super key*" (45%), "keuntungan/*benefit*" (40%), "foreign key/*foreign key*" (35%), "candidate key/*candidate key*" (35%), "sederhana/*simple*" (35%), "baris dan kolom/*rows and fields*" (35%), "kolom/*field*" (35%), "primary key/*primary key*" (35%), "alternate key/*alternate key*" (30%), and "baris/*rows*" (30%).

Further analysis was carried out to reveal frequent relationship patterns that expressed learners' understanding based on the propositions they made. Table 1. shows the frequent itemset of relationships with minimum support of 30%. The pattern of frequent relationships looks constant with the findings of the frequent itemset of concepts.

Table 1. *Frequent itemset of relationships*

Relationships	**Frequency**	**Support (%)**
basis data relasional – tabel dua dimensi	10	50
basis data relasional – istilah	8	40
istilah – atribut	8	40
istilah – relasi	8	40
istilah – tupel	8	40
istilah – domain	7	35
basis data relasional – keuntungan	7	35
istilah – derajat	6	30
istilah – cardinality	6	30

The frequent sub-concept map analysis was conducted to reveal the possibility of the appearance of new hidden patterns. Frequent sub-concept maps explain the association of proposition a and proposition b on concept maps. By setting minimum support of 30%, three sub-concept map patterns consisting of 2 propositions were obtained. The association {basis data – degree, basis data – domain} having a support value of 35%, {basis data – derajat, basis data – relasi} was 30%, and {basis data - domain, basis data – relasi} was 30%.

Further analysis was carried out to find association rules that were formed by involving the minimum confidence requirements. By setting the 30% support threshold, association rules were obtained, as shown in Table 2. Six association rules defined the best rules with a confidence value of 100% and support of 40%. Furthermore, the second-best rule with a confidence value of 87.5% and support of 35% formed nine association rules. The following best rule produced seven association rules with a confidence value of 85.71% and support of 30%. Associative rules that were formed provide new information for teachers for decision-making.

Table 2. *Association rules of propositions*

Association rules	**Support (%)**	**Confidence (%)**
istilah – atribut => istilah – relasi	40	100
istilah – atribut => istilah – tupel	40	100
istilah – relasi => istilah – atribut	40	100

istilah – relasi => istilah – tupel	40	100
istilah – tupel => istilah – relasi	40	100
istilah – tupel => istilah – atribut	40	100
istilah – atribut => istilah – domain	35	87.5
istilah – atribut => basis data relasional – istilah	35	87.5
istilah – relasi => istilah – domain	35	87.5
istilah – relasi => basis data relasional – istilah	35	87.5
istilah – tupel => istilah – domain	35	87.5
istilah – tupel => basis data relasional – istilah	35	87.5
basis data relasional – istilah => istilah – atribut	35	87.5
basis data relasional – istilah => istilah – tupel	35	87.5
basis data relasional – istilah => istilah – relasi	35	87.5
istilah – domain => istilah – derajat	30	85.71
istilah – domain => istilah – cardinality	30	85.71
istilah – domain => basis data relasional – istilah	30	85.71
basis data relasional – keuntungan => istilah – relasi	30	85.71
basis data relasional – keuntungan => istilah – atribut	30	85.71
basis data relasional – keuntungan => istilah – tupel	30	85.71
basis data relasional – keuntungan => basis data relasional – istilah	30	85.71

In line with previous research, data mining techniques have the potential to be beneficial for uncovering hidden information in the educational context (Yoo and Cho, 2012; Adeitan and Salau, 2019). This study found that data mining techniques could reveal students' understanding of an open-ended concept map. In particular, teachers could quickly find out frequent concepts, frequent relationships, and association rules formed by learners through the operation of association rules.

The teacher received many benefits after revealing the frequent itemset of relationships, as it emphasized that relationships are representations of propositions on the concept map (Prasetya et al., 2021). Each occurrence of a proposition was given a weight to reveal its role in the concept map. Finally, through the forming of association rules, teachers could capture the understanding patterns of their students. This information has an important meaning in revealing what they think after gaining new knowledge.

4. Conclusion and Future Work

The present study utilized data mining techniques to find hidden information on open-ended concept maps. This initial study found that data mining techniques have the positive potential for interpreting large sets of open-ended concept maps. The analysis results suggested that the learners' understanding could be revealed through frequent itemset of concepts, frequent itemset of relationships, and patterns of association rules that are formed. Compared to manual evaluation, which is time-consuming and challenging, association analysis finds valuable information shortly. Furthermore, the obtained associative rules could be used to identify the extent to which the learners understand the material that the teacher has delivered.

The present study has several limitations that should be considered. First, the number of concept maps as data sets involved in this experiment was relatively small and needed to be increased to achieve more reliable results. Second, this study employed the Apriori algorithm, which tends to be less efficient in handling large data sets. Therefore, other algorithms such as FP-Growth, Generalized Rules Induction, or hash-based can be tested. In addition, the present study focused on association rules which simply reveal the percentage of proposition combinations. Other operations such as similarity analysis of learners' concept maps and teacher's maps are fascinating to be applied.

Acknowledgments

The first author would like to acknowledge the Universitas Negeri Malang (UM) for supporting this research.

References

Adekitan A. I & Salau, O. (2019). The impact of engineering students' performance in the first three years on their graduation result using educational data mining, Heliyon, 5(2), e01250, 2019.

Ait-Mlouk, F. Gharnati, and T. Agouti. (2017). An improved approach for association rule mining using a multi-criteria decision support system: a case study in road safety. European transport research review, 9(3), 1-13.

Alasadi and W. S. Bhaya, Review of data preprocessing techniques in data mining. Journal of Engineering and Applied Sciences, 12 (16), 4102-4107, 2017.

Aleven, V., McLaren, B. M., Sewall, J., Van Velsen, M., Popescu, O., Demi, S. & Koedinger, K. R. (2016). Example-tracing tutors: Intelligent tutor development for non-programmers, International Journal of Artificial Intelligence in Education, 26(1), pp. 224-269.

Angiani, L. Ferrari, T. Fontanini, P. Fornacciari, E. Iotti, F. Magliani, and S. Manicardi. (2016). A Comparison between Preprocessing Techniques for Sentiment Analysis in Twitter. In KDWeb, 2016.

Baker. (2019). Challenges for the future of educational data mining: The Baker learning analytics prizes. Journal of Educational DataMining,11(1), 1–17.

Cañas, A. J & Novak, J. D. (2014). Concept mapping using CmapTools to enhance meaningful learning, In Knowledge cartography, pp. 23-45. Springer, London.

Hirashima, T. (2019). Reconstructional concept map: automatic Assessment and reciprocal reconstruction, International Journal of Innovation, Creativity and Change, Vol. 5, pp. 669-682.

Karabatak, M. C. Ince, and A. Sengur. (2017). Wavelet domain association rules for efficient texture classification. Applied Soft Computing, 11(1), 32-38, 2011.

Pailai, J., W. Wunnasri, K. Yoshida, Y. Hayashi, and T. Hirashima. (2017). The practical use of Kit-Build concept map on formative assessment. Research and Practice in Technology Enhanced Learning, 12(1), pp. 1-23, 2017.

Pinandito, A., Prasetya, D., Hayashi, Y and Hirashima, T. (2021). Design and development of semi-automatic concept map authoring support tool. Research and Practice in Technology Enhanced Learning, 16(1), 1-19.

Prasetya, D. D., T. Widiyaningtyas, T. Hirashima, and Y. Hayashi. (2019). Reconstruction of Concept Map to Promote Learners' Comprehension on New Knowledge, In The 27th International Conference on Computers in Education (ICCE 2019), volume 1, pages 369-371, 2019.

Prasetya, D. D., T. Hirashima, and Y. Hayashi. (2021) Comparing Two Extended Concept Mapping Approaches to Investigate the Distribution of Students' Achievements, IEICE TRANSACTIONS on Information and System. Vol. 104 (2), pp. 337-340.

Romero and S. Ventura. (2020). Educational data mining and learning analytics: An updated survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1355.

Shao, Y. Li, X. Wang, X. Zhao, and Y. Guo. (2020). Research on a new automatic generation algorithm of concept map based on text analysis and association rules mining. Journal of ambient intelligence and humanized computing, 11 (2), 539-551.

Taricani, E. M. and R. B. Clariana. (2006). A technique for automatically scoring open-ended concept maps, Educational Technology Research & Development, 54: pp.65-82.

Villalon and Calvo, R. A. (2008). Concept map mining: A definition and a framework for its evaluation. In 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (Vol. 3, pp. 357-360), IEEE.

Xu, Y. (2016). Research of Association Rules Algorithm in Data Mining, International Journal of Database Theory and Application 2016; Jun 30;9(6): 119-30, 2016.

Yoo, J.S. & M. H. Cho. (2012). Mining Concept Maps to Understand University Students' Learning. International Educational Data Mining Society.