Data Mining: An Overview and Application of Association Rule Mining

ICTU Student 10
Thai Nguyen University of Information and Communication Technology
Thai Nguyen University, Thai Nguyen, Vietnam

***Abstract*—This thesis provides an in-depth overview of data mining, including its definition, types of data, and the process of data mining. It focuses on association rule mining, detailing various algorithms such as Apriori and AprioriTid. The author applies these algorithms to a credit course registration database, identifying relevant patterns and rules. The work highlights the ongoing development and improvement of data mining algorithms, particularly in association rule mining, and discusses their practical application in educational settings, such as credit course registration databases. Despite limitations in time and resources, the author aims to contribute to the field and improve future research.**

***Index Terms*—data mining, association rules, Apriori, AprioriTid, credit course registration, database**

I. Introduction

The thesis introduces data mining as a field of study with significant practical applications. It covers fundamental concepts, types of data, and the data mining process, emphasizing the importance and development of association rule mining algorithms. The author discusses the significance of data mining in various domains, including education, and highlights the need for efficient and effective data analysis techniques. The introduction also outlines the structure of the thesis, which includes an overview of data mining, a detailed discussion of association rule mining algorithms, and an application of these algorithms to a practical scenario. [1]

II. Methods

The author discusses several association rule mining algorithms, including Apriori, AprioriTid, and others. These algorithms are applied to a credit course registration database to identify useful patterns and rules. The thesis provides a detailed explanation of the Apriori algorithm, which involves generating frequent itemsets and then using these to generate association rules. The AprioriTid algorithm is also discussed, which is an extension of the Apriori algorithm that incorporates transaction IDs to improve efficiency. The author explains the steps involved in applying these algorithms, such as setting the minimum support and confidence thresholds, and the process of generating and evaluating association rules. [1]

III. Results

The application of association rule mining algorithms to the credit course registration database resulted in the identification of relevant patterns and rules, though specific performance metrics are not provided. The author identifies several frequent itemsets and association rules that could be useful for understanding student behavior and course preferences. For example, the thesis reports that students who register for mathematics courses are more likely to also register for statistics courses. However, the author notes that the results are preliminary and further validation is needed to ensure the reliability of the identified patterns. [1]

Supervisor: ICTU Teacher 4, Thai Nguyen University of Information and Communication Technology.

IV. Discussion

The thesis highlights the ongoing development and improvement of data mining algorithms, particularly in association rule mining. It discusses the practical application of these algorithms in educational settings, such as credit course registration databases, to support decision-making processes. The author emphasizes the importance of selecting appropriate algorithms based on the specific requirements of the application. The discussion also addresses the limitations of the current work, such as the limited time and resources available, and the need for more comprehensive evaluation methods. The author suggests that future research should focus on improving the efficiency and accuracy of association rule mining algorithms and their integration into real-world applications. [1]

V. Conclusion

Data mining is a relatively new and promising field in information technology. The thesis presents an overview of data mining and delves into association rule mining algorithms, applying them to a practical scenario. Despite limitations in time and resources, the author aims to contribute to the field and improve future research. The author acknowledges the need for further validation and comprehensive evaluation methods to ensure the reliability of the identified patterns and rules. The work highlights the potential of data mining in supporting decision-making processes in educational settings and other domains. [1]

References

[1] ICTU Student 10, ”Data Mining: An Overview and Application of Association Rule Mining,” undergraduate thesis (original in Vietnamese), Thai Nguyen University of Information and Communication Technology, n.d.

[2] www.cs.waikato.ac.nz/ml/weka

[3] www.sgi.com/tech/mlc/ - 25k

[4] www.visual-basic-data-mining.net/ forum/ShowForum.aspx

[5] dmoz.org/Computers/Software/Databases/Data_Mining/Public_Domain _Software

[6] www.planet-source-code.com/vb/ scripts/ShowCode.asp

[7] www.cs.concordia.ca/db/dbdm/dm.html