ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20230327000481
Original scientific paper

An Improved Apriori Algorithm for Association Rule Mining in Employability Analysis

Fang PENG, Yuhui SUN*, Zigen CHEN, Jing GAO

Abstract: The wide application of emerging advanced technologies causes significant changes in the development trend of the employment market. The lack of flexible and easy-to-implement analysis methods challenges general maritime education practitioners to understand the developing trends. This research proposed the improved Apriori algorithm to explore employment preference by identifying the association rule of the employability indicators and the employment status. The candidate generation methods are optimised based on the questionnaire design to generate fewer candidates. The minimum support value is automatically generated to reduce the reliance on analysis expertise and improve accuracy. To validate the algorithm, a questionnaire for the maritime graduate is used to collect employment data to test the efficiency and capability of the improved algorithm. The computation time for different data set sizes shows that the improvement could improve the algorithm's effectiveness. The algorithm also successfully identifies significant employment preference that certain employment types emphasise specific employability skills, such as responsibility and core professional skills. The results suggest that the improved A algorithm could reduce the computing burden and identify the employment preference from questionnaire data. This research provides easy-to-use and flexible analysis tools, which could reduce the computing expertise required for education practitioners.

Keywords: apriori algorithm; association analysis; education analysis; employability

1 INTRODUCTION

The maritime industry involves the core logistics operation for cargo transportation worldwide, and its international influence has increased dramatically in the past few decades [1]. As an essential part of the global industry, the maritime industry transports about 90% of international trade products [2]. With the accelerated application of advanced technologies for maritime market competition, the proportion of international trade freight volume will continue to increase [3]. The vigorous development of the maritime industry is inseparable from the firm guarantee of adequate maritime talent supply. However, the traditional knowledge system and employability system of maritime industry practitioners will face upgrading and transformation. Technological revolutions have taken place in ships, electronic systems, and automatic control systems, which require a comprehensive revision of maritime education and training requirements to meet the actual industrial requirements [4]. The strict safety requirements also require employees to solve the problems of ship operation on-site for the long duration of the voyage [4]. Therefore, the maritime talent training system that can adapt to future employment demands needs to be restructured accordingly. Effectively cultivating comprehensive and high-quality maritime talent is an important and meaningful research issue.

Overcoming such difficulties and challenges from the development of the maritime industry requires an adequate and comprehensive assessment of the current maritime employment status and employability skills. Current maritime employability studies commonly focus on analysing the significance of employability skills in different scenarios [5, 6]. Those studies lack an analysis of the correlation between the employability skillset and the general status of the employment market. The association rule mining has been used to assess education; however, the main research focuses are on the teaching methods and the improvement of student performance and attitude [7, 8]. Many scholars have researched employability, but there is a lack of in-depth research in the maritime field [9]. Association rule mining is essential for exploring the relationship in large datasets across many research domains, including education analysis [10]. The algorithm for association rule mining could identify the strong rule hidden in a large quantity of employment data, which is beneficial to maritime employability research.

The Apriori algorithm is one of the commonly used algorithms for association rule mining [11]. The algorithm first identifies all frequent individual items in a dataset and combines these frequent items to generate larger frequent item sets. Then the algorithm iteratively identifies and combines the larger item sets until no more frequent ones can be found. During the process, a minimum support threshold is used as a criterion to select the frequent item [12]. The Apriori algorithm has the advantage of easy implementation, intuitive interpretation, simplicity and scalability for efficiently handling large datasets [13]. However, the iteration steps generate a large number of candidate item sets which are mostly infrequent, so the algorithm can be ineffective in such senario. The Apriori algorithm has great potential in finding employment preference as the association rule. However, since the employability analysis involves many maritime employers and employees, the algorithm can be computationally expensive [5]. Therefore, the Apriori algorithm requires some improvement to analyse a large amount of the employability questionnaire data.

This study aims to identify employability skills by analysing the employment questionnaire data with the improved Apriori algorithm. The improved Apriori algorithm proposed by this study has simplified the candidate generation method based on the structure of the questionnaire to improve the algorithm efficiency. This study also designed the self-adjust minimum support generation method to improve the accuracy of selecting the association rules. With the improved Apriori algorithm, this study has identified the association rule between employability skills and certain employment types. The identified association rules are further analysed to explore the employment pattern of the current employability demands. The contribution of the research is two-fold: the first contribution is that the proposed Apriori algorithm could improve the employability analysis efficiency; the second contribution is that the association rules guide the development of the student's employability towards the requirement of the employment market. Following this introduction, Section 2 reviews the related concepts of associated rule mining and apriori algorithm; the proposed Apriori algorithm for questionnaire analysis is presented in section 3; Section 4 outlines questionnaire design and data collection as the context of the data analysis; Section5 presents the detailed analysis process and the identified association rules; Section 6 discusses the employability demands based on the identified rule and the limitation of the algorithm. The last section concludes this study.

2 LITERATURE REVIEW

2.1 Development of Maritime Industry and Employment Challenge

The transportation industry innovation is promoted by the new technology advancement from related industries. The development direction of the maritime industry includes intelligence and automation, environmental protection and sustainable development, and industry optimisation and integration. The development trend of the maritime industry and market globalisation has greatly affected the market's requirements for the employability of maritime graduates [14]. Intelligence and automation are manifested in the intelligence of the shipping logistics supply chain and the automation of ships and ports [15]. The smart ship has significant advantages in ensuring navigation safety, improving ship energy efficiency, and reducing transportation costs [16, 17]. With the continuous increase in electrification, digitalisation, and integration, new problems and new challenges have been brought to the maritime industry personnel [18].

For environmental protection and sustainable development, the current development trend has three main methods to reduce marine greenhouse gas emissions: using green technologies to improve sustainability, using market policy-based adjustment tools, and using effectively optimising shipping planning [19]. In addition, there is research on using port power supply facilities to provide power to docked ships [20]. Reducing the navigation speed of ships in the coastal waters can also effectively reduce greenhouse gas and harmful gas emissions [21, 22]. The emphasis on environmental protection requirements, corresponding laws, and regulations also requires seafarers to improve their awareness and skills in sustainability measures.

Another development direction of the maritime industry is the integration of the maritime-related industry [23]. Combining a series of transportation methods, such as trains, aeroplanes, and road transportation, the integrated maritime industry has many advantages [24]. The vertical integration of the shipping industry brings valuable options for differentiating from competitors. Since their transportation time, pricing, and reliability are roughly the same, the differences in the services offer significant competitive advantages [25]. These integration trends in the global supply chain have forced maritime talent to adopt the business aspect knowledge in addition to new generations of computer information technology knowledge.

2.2 Associated Rule Mining for Education Assessment

Association rule mining is the analysis method to identify the association and the correlations between the large volume of data sets [10]. The identified association could reflect the laws and patterns in the appearance of the object attributes.

Two criteria commonly used for association rule mining are support and confidence. The degree of support and confidence determines the reliability and certainty of the rules identified in the mining process [26]. The support criteria represent the frequency of the item in the dataset, whereas the confidence criteria represent the correctness of these association rules [12]. The association rules that meet the minimum requirement of support and confidence simultaneously are the strong rules. Association rule mining has wide applications in several domains, including education assessment. With the advancement of education digitalisation, considerable information and data on student performance could be collected for further analysis [27]. Those datasets could contain certain teaching rules and learning patterns, which could be uncovered by association rule mining. The identified rules could assess the status of the teaching and learning process and guide the improvement of the education practice [27]. Thus the association rule mining for education has received much attention from scholars. Alangari and Alturki [28] use the data mining technique to assess performance and its association with student attributes such as course grades. García, Romero [8] use association rule mining to analyse the web-based course data to obtain the association between student usage and the course information, thus establishing the course recommendation structure. Hashima, Hamoud [29] adopt association rule mining to discover the factors impacting the student's success to provide guidance on improving the student performance.

2.3 Apriori Algorithm

The Apriori algorithm is commonly used for association rule mining [11]. The core principle of the algorithm is that if one itemset is the frequent itemset, then all its subsets are the frequent itemset; if one itemset is not the frequent itemset, then all its subsets are not the frequent itemset. The algorithm identifies the association rule by finding all the frequent itemsets in the data and then exploring the association rules from the frequent dataset [12]. However, the Apriori algorithm also has several limitations. The algorithm could be slow with a large volume of the dataset, a high dimensional itemset, or a low support value. The large data volume could generate a large initial data set; the low support value could result in a more frequent itemset for each iteration, and the high dimensional dataset could result in significantly more iterations [13, 30]. Those factors could significantly increase the calculation time and decrease the algorithm efficiency. Also, with many frequent itemsets and transactions, the algorithm requires significant computing resources such as CPU and memory [31]. Many analysis methods have been developed to provide more accurate, efficient employability analysis, including Stream mining approaches, Multi-criteria association rule mining, and Hybrid approaches. Stream mining approaches can handle large volumes of data and provide real-time insights [32]. Multi-criteria association rule mining can provide more nuanced and actionable insights by considering multiple criteria simultaneously [33]. Hybrid approaches can handle complex relationships and data types, which are challenging for the Apriori algorithm [34]. Although those analysis methods provide a considerable advantage, the Apriori algorithm still has the benefit of easy implementation and intuitive interpretation. For employment research such as employment preference, the flexibility in designing the experiment is important. Therefore, the Apriori algorithm is still important for educational research to design the questionnaire based experiments.

3 RESEARCH METHOD

This study first proposes the Apriori algorithm with improved candidate generation and the automatically generated minimal support value method. Then this study used maritime employment questionnaire data to test the algorithm's efficiency and whether the algorithm could identify employment preference. The proposed algorithm has two improvements. The first improvement is the optimised candidate generation method which only generates candidates relevant to employment preference. The optimisation is based on the common design of the employment questionnaire. The second improvement is the automatically generated minimal support value from the mean and standard deviation of the questionnaire data. For the validation test, this research conducts a large-scale survey to explore the character of employability indicators. The design of the questionnaire and the data pre-processing are discussed in detail because the candidate generation method is based on it. The computation time of the original Apriori algorithm and the proposed Apriori algorithm is compared for the algorithm's efficiency. The efficiency is tested with different sizes of questionnaire data. This study also analyses all the questionnaire data to identify the association rule of the employability indicators.

3.1 Apriori Algorithm Design

The essential process of the Apriori algorithm is about calculating the support and confidence values. The frequent set is identified by finding all the items meeting the support and confidence threshold which are set by the algorithm user. For different items (X and Y), support for item X is the ratio of item X in all itemsets, as Eq. (1):

$$Support(X,Y) = P(X,Y) = \frac{number(X,Y)}{number(\text{All sample})} \tag{1}$$

The confidence of the generated association rules from the frequent itemset is the ratio of the number of occurrences of items containing both X and Y to the number of occurrences of items containing X, as Eq. (2):

$$Confidence(X,Y) = P\left(\frac{Y}{X}\right) = \frac{P(X,Y)}{P(X)} = \frac{number(X,Y)}{number(X)} \tag{2}$$

The basic process of the algorithm is:

- Identify all the items which fit the parameter $k = 1$, conduct dataset scanning to determine the support value for each $k = 1$ item.
- Identify the $k = 1$ itemset with support higher than the minimum support as the frequent item set to obtain the frequent set $S(k = 1)$.
- Generating new candidate k itemset using the frequent set $S(k - 1)$ generated from the last iteration.
- Rescanning the dataset to determine the support value for each k itemset candidate. Identify the k itemset candidate with a support value higher than the minimum support, and form the frequent set $S(k)$.
- Continue the iteration until the algorithm does not generate a new frequent itemset.

The algorithm is demonstrated in the flowchart as Fig. 1.

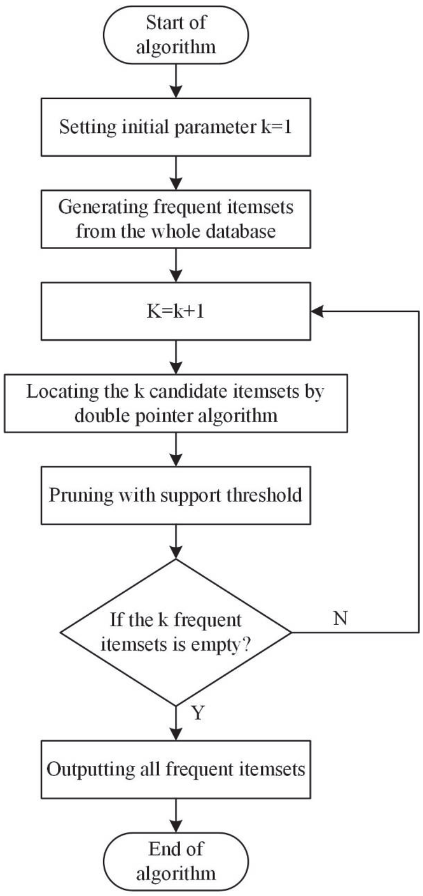

Figure 1 The flow diagram of the Apriori algorithm

3.2 Improved Candidate Generation Method

One of the significant limitations of the Apriori algorithm is that the generation of a large number of itemsets leads to a significant computation burden of scanning the database to calculate the support value for each generated itemset. Therefore, limiting the itemset generation is necessary for algorithm improvement. This research proposes a new approach to generate the item set according to the questionnaire setting and the research objective of this research. This research intends to study the employability skills highly demanded by different types of employment enterprises. Therefore, the itemset generating has the following rules: From the second iteration, the itemset candidate should only contain one item from the employability indicator itemset "I" and contain one or several items from the background variable itemset "B". For the latter iteration, due to the limited size of the background set "B", each itemset generated in the iteration produces fewer itemsets than the original algorithm.

- Generate the itemset candidate with the standard itemset generation method.
- Check if the intersection between the itemset and the background set "B" is not empty and if the intersection between the itemset and the indicator set "I" has only one element.
- Output the filtered itemset.

3.3 The Automatic Support Threshold Generation Method

The Apriori algorithm requires the users to set up the minimum support value and the minimum confidence value as the criteria to identify the frequent itemset and the strong association rules. Selecting those criteria thresholds requires experience in algorithm implementation and in-depth knowledge of the study subject. However, for many researchers in the education field, determining suitable criteria thresholds could pose a challenge. Thus, this research proposes auto-generating the minimum support value based on the attribute of the datasets to effectively and adequately identify the frequent itemsets. The minimum support is calculated by calculating the mean of each item's support plus the standard deviation of the support values. The mean and standard deviation is calculated as Eq. (3) and Eq. (4). The support [i] is the support value of item i. N is the total number of items.

$$Mean = \frac{\sum_{i=1}^{N} Support[i]}{N} \tag{3}$$

$$Standard\ deviation = \sqrt{\frac{\sum_{i=1}^{N} \left(Support[i] - Mean\right)^2}{N-1}} \tag{4}$$

Both the new itemset candidate generation and the self-generated minimum support values are used for each Apriori algorithm iteration. Since the modifications of the Apriori algorithm only concern the itemset candidates generation and the criteria thresholds, the main workflow of the algorithm remains unchanged. Therefore, the implementation of the algorithm is similar to the algorithm flow introduced in Fig. 1.

3.4 The Questionnaire Design and Data Pre-Processing Method

The questionnaire is based on multiple questionnaires for student capability measurement from Dalian maritime university and the suggestion from the experts of the Delphi study. The structure of the questionnaire influences the candidate generation method proposed in this paper. Therefore, detailed research aims and an understanding of the questionnaire are essential for the algorithm implementation. The questionnaire contains two parts. The first part is the participants' background information, and the second part requires respondents to assess and score the importance of various employability indicators. There are 14 employability skills which are listed in Tab. 1. The questionnaire adopts the Likert 5-point scale as the test method to score the corresponding employability indicators.

Table 1 Initially employability indicator system for maritime graduates

Primary indicator	Secondary indicator	Index
Personal qualities	Adaptability	I1
	Learning and self-development capability	I2
	Critical thinking	I3
	Responsibility	I4
Foundational capability	General management capability	I5
	Language skills	I6
	Teamwork and communication capability	I7
	Problem-solving capability	I8
Professional capability	Nautical competence	I9
	Equipment operation and maintenance capability	I10
	ICT skills	I11
	Cargo management capability	I12
	Maritime business skill	I13
	Implementation of international conventions	I14

The questionnaire data is in the format of the 5-point scale rating, which is not suitable for the algorithm. Thus, the questionnaire data is formatted according to the algorithm design. The core principle of formatting is about transforming the numerical data into text data with the indication of the question index. There are two sections of the questionnaire: one for background questions and one for employability indicators. For the answers to background questions, the letter "B", the question number and the letter for the answer are used to form the text data (For example, B1-A). For the employability question, the 5-point scale is interpreted with the following methods. The scores of 4 and 5 are regarded as important for the corresponding employment type and marked with the text "I" in the formatting; the score of 3 is regarded as neutral and marked with "N"; the scores of 1 and 2 are regarded as not important and marked as "NI". Therefore, the letter, the question number, and the text label are used to form the text data (for example, I1-important). The questionnaire data are exported and formatted with excel. A sample for the data formatting is demonstrated in Fig. 2.

Original questionnaire data example

Participants	Background question				Indicator question section				
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	…	Q14
1	C	A	A	B	5	3	4	…	3
2	A	B	C	A	4	2	1	…	4

Dataset after formatting example

Participants	Background item group				Indicator item group				
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	…	Q14
1	B1-C	B2-A	B3-A	B4-B	I1-I	I2-N	I3-I	…	I14-N
2	B1-A	B2-B	B3-C	B4-A	I1-I	I2-NI	I3-NI	…	I14-I

Figure 2 The example of the formatting of the questionnaire data

3.5 Experiment Setting

This research conducts a large-scale survey to explore the character of employability indicators and use the Apriori algorithm to identify the association rules. This paper selects a wide range of companies hiring maritime graduates recruitment for the large-scale survey, including state-owned companies, private companies, seafarer dispatch agencies, and other types of companies. These companies or agencies have engaged in a wide range of maritime operations, including crew management, crew dispatch, seafarer recruitment and placement, ship transportation, ship trading, and shipping agency. Moreover, the company's distribution area covers most of the Chinese maritime industry, including the centre of the maritime industry, like Shanghai, Beijing, Guangzhou, Shenzhen, and other coastal cities, and related areas like Taiwan, Hong Kong, and Singapore. The questionnaires are distributed through the alumni association, employment, and other related organisations of Dalian Maritime University. Personals are invited to participate in the online survey anonymously. The questionnaire survey started in August 2019 and ended on October 31, 2019; a total of 2354 questionnaires were received. After the questionnaires were retrieved, preliminary sorting was carried out to eliminate 742 invalid questionnaires. There were 1612 qualified questionnaires with an effective questionnaire recovery rate of 68.48%.

The experiment is conducted on a laptop computer with Intel i5 CPU and an SSD hard drive. For the efficiency test, this study uses five groups of questionnaire data with data sizes of 500, 1000, 1500, 2000, and 2500. Since the original questionnaire data only have 1612 qualified questionnaires, the rest are generated from the original questionnaire data by randomly changing 25% of the answers. The computation time of the original Apriori algorithm and the proposed Apriori algorithm for five datasets are documented to compare the difference. For the employability indicator study, all 1612 qualified questionnaire data are analysed with the proposed Apriori algorithm to find the frequent item set and the association rules. The algorithm parameter setting for the association rule mining is quite simple. Since the minimal support value is generated by the algorithm, the only required parameter is the minimal confidence level which is set at 75%. The setting of the minimal confidence level is relatively low due to the expectation that the maritime employment market is diverse and does not have dominating employment preferences.

4 RESULTS
4.1 The Algorithm Efficiency Comparison

The results from the validation test suggest that the improved Apriori algorithm has higher efficiency than the traditional Apriori algorithm. The improved Apriori algorithm was compared with the original algorithm in terms of performance. Fig. 3 contrasts the execution time of the two algorithms under different data set sizes. The improvement in computation efficiency is mainly due to the improvement of the candidate generation methods. In each iteration, the improved algorithm only generates the frequent item sets related to the employment preference; thus, the frequent sets with employment information and employability indicator at the same time will be generated. The following process of finding and comparing the frequent item also requires less calculation. The improved algorithm could save more computation resources on larger datasets. The proposed algorithm used 70% of the computation time at 500 datasets, whereas the proposed algorithm only used 57% at 2500 datasets compared with the traditional algorithm. Therefore, the proposed algorithm is more suitable for analysing the large quantity of employment data at the national maritime employment market level.

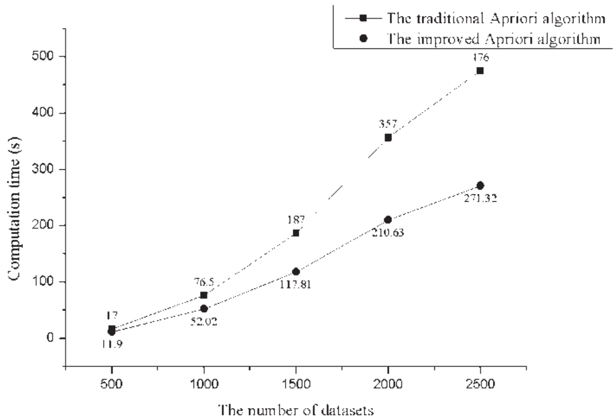

Figure 3 The algorithm execution time comparison

4.2 Identified Frequent Items and Rule Generation

After the data formatting, the questionnaire data is processed with the improved Apriori algorithm. The support value threshold is automatically generated with the designed formula and outputted by the algorithm. The minimum support value generated is 0.54025. With the support value threshold of 0.54025, the improved Apriori algorithm output a total number of 7 frequent items, which are listed in Tab. 2.

Table 2 The frequent itemset identified from the improved algorithm

No	Frequent items	Support
1	(B1_1, I9_important)	0.55087
2	(B3_1, I4_important)	0.54156
3	(B3_1, I9_important)	0.54218
4	(B3_1, I10_important)	0.54777
5	(B4_1, I4_important)	0.55769
6	(B4_1, I9_important)	0.56638
7	(B4_1, I13_not_important)	0.54404

Then the association rules are generated from the frequent itemset. The association rules with a confidence level of over 0.7 are listed in Tab. 3.

Table 3 The association rules generated from the frequent itemset

No	Antecedent	Consequent	Attitude	Support	Confidence
1	Nautical Science	Nautical competence	important	0.5508685	0.8671875
2	State-owned enterprise	Responsibility	important	0.5415633	0.75649913
3	State-owned enterprise	Nautical competence	important	0.5421836	0.75736568
4	State-owned enterprise	Equipment operation and maintenance capability	important	0.5477667	0.76516464
5	Large enterprises	Responsibility	important	0.5576923	0.78105995
6	Large enterprises	Nautical competence	important	0.5663772	0.79322328
7	Large enterprises	Maritime business skill	not important	0.5440447	0.76194613

All association rule from the identified frequent itemsets has a high confidence level with a minimum of 0.75. The high confidence level suggests that the questionnaire data have strong association rules. The identified association rules are mainly related to the employment types of state-owned enterprises and large enterprises.

5 DISCUSSION

5.1 The Demands of Employability Indicators from Different Employment Types

From Tab. 3, the state-owned enterprises favour the employability skills of responsibility, nautical competence, and equipment operation and maintenance capability. Most enterprises value responsibility as maritime logistic operations have the characteristics of long operation duration, harsh operating environment, and high reliance on machinery. Therefore, the association rule related to responsibility is the expected outcome of the questionnaire analysis. The other two indicators are the professional skillset of the maritime major associated with nautical navigation and the equipment operating. Those are the essential aspects of the maritime operation.

For the large size enterprise, the identified association rules are similar to the state-owned enterprise, containing the responsibility and the professional skillset. These rules suggest that the core requirements of different enterprise types in the maritime employment market are quite similar. However, large enterprises also have one interesting rule: large companies do not value maritime business skills for graduate employment. As the maritime industry becomes integrated, the general thinking is that business knowledge could be the basic requirement to promote stable and efficient operations. Thus, business skills could be valuable for some enterprises. But the large enterprise could have significantly more employees, which means that the large enterprise could afford specific personnel to handle the business routine. Therefore, the requirement for graduates in maritime business skills is reduced.

The Apriori did not produce important association rules for the other two background information types. The algorithm detected that the employment type of nautical science emphasises professional nautical competence. Such an outcome is expected as it is a basic professional skill for that employment type. The location of the maritime enterprise does not appear in the frequent itemsets and the association rules, which means that the enterprise location does not lead to any specific requirement of the employment market. However, as more and more people from inland provinces join marine enterprises, this background factor still deserves attention.

The algorithm proves to be a valuable tool for analysing a large quantity of questionnaire data and providing guidance for the educator and employer. The improved Apriori algorithm can quickly and objectively analyse large datasets and generate association rules for informed decisions. The education practitioner could use the algorithm to identify the employment preference to improve education in maritime occupations with new requirements for advanced technologies. They can ensure that education and training programs are designed to meet the job market's needs. Employers can also use the Apriori algorithm to identify the skills and qualifications that are required for career progression within their organisations. With a better understanding of the employment market, the employer can set suitable recruitment strategies and job descriptions to attract the best candidates and provide training and development opportunities for their employees

5.2 The Limitation of Applying Apriori Algorithm to Questionnaire Data

One limitation of the Apriori algorithm in this questionnaire data analysis is that the algorithm failed to generate certain expected outcomes. The Apriori algorithm only identifies the common association relations and ignores the less common and important relationship in the datasets. The lack of those expected rules could be explained by the questionnaire participants not being evenly distributed across the background types. For example, most employment in the Chinese maritime market is in the catalogue of state-owned enterprises, which accounts for over 75% of the participants. The participants from enterprise dispatching labour abroad only accounts for about 12%. Thus, the item "B3-3" will not be identified as frequent, and all association rules regarding the enterprise dispatching labour abroad will not be detected with the current analysis setting.

Another limitation of the Apriori algorithm is that algorithm could not handle missing data and deal with noisy or incomplete datasets. The employability data from the questionnaire are likely to have missing question items and incorrectly filled answers. The missing data will cause the algorithm to provide incorrect and biased results. For example, if the questionnaire data contain many unanswered questions, the algorithm will give biased and incorrect association rules against the employee with that background. This limitation requires a considerable data preparation process before applying the algorithm.

Further research directions could be optimising the generating rules of minimum support value for more accurate rule detection and improving pattern matching method to reduce total database scan over the whole calculation. Including domain expertise and trial-and-error testing would also help improve the minimum support value generation. Other research directions could be improving the missing data and noise handling. The techniques such as the data augmentation and feature selection could be integrated into the Apriori algorithm to improve the accuracy of the association rule mining results.

6 CONCLUSION

This research proposes an improved Apriori algorithm to study the employability skill demands of different employment types. The results show that the proposed Apriori algorithm could effectively identify the association rule from the educational questionnaire data. However, the algorithm has the issue of not identifying the association relationship in the minor data category. Therefore, the application of Apriori algorithm has the potential in questionnaire data analysis but requires further improvement. The association relationship finding of this research could help improve the employability cultivation of the maritime graduate to meet the need of different employment types. Further research directions could be optimising the generating rules of minimum support value for more accurate rule detection and improving pattern matching method to reduce total database scan over the whole calculation.

8 REFERENCES

[1] DeSombre, E. R. (2006). *Flagging standards: globalization and environmental, safety, and labor regulations at sea*. MIT Press Books, 1.

[2] Alderton, T. & Winchester, N. (2002). Globalisation and de-regulation in the maritime industry. *Marine Policy*, *26*(1), 35-43. https://doi.org/10.1016/S0308-597X(01)00034-3

[3] Syriopoulos, T. & Theotokas, I. (2007). Value creation through corporate destruction? Corporate governance in shipping takeovers. *Maritime Policy & Management*, *34*(3), 225-242. https://doi.org/10.1080/03088830701342973

[4] Nasaruddin, M. M. & Emad, G. R. (2019). Preparing maritime professionals for their future roles in a digitalized era: bridging the blockchain skills gap in maritime education and training. *Proceedings of the International Association of Maritime Universities (IAMU) Conference.*

[5] Kabir, M. (2014). *Enhancement of seafarers' employability through capacity building in maritime education and training (MET): a case study of Bangladesh.* Master Degree, World Maritime University Dissertations.

[6] Chen, P. S.-L., Cahoon, S., Pateman, H., Bhaskar, P., Wang, G., & Parsons, J. (2018). Employability skills of maritime business graduates: industry perspectives. *WMU Journal of Maritime Affairs*, *17*(2), 267-292. https://doi.org/10.1007/s13437-018-0140-9

[7] Verma, S. K. & Thakur, R. (2017). Fuzzy Association Rule Mining based Model to Predict Students' Performance. *International Journal of Electrical & Computer Engineering (2088-8708)*, *7*(4). https://doi.org/10.11591/ijece.v7i4.pp2223-2231

[8] García, E., Romero, C., Ventura, S., & Castro, C. d. (2009). An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering. *User Modeling and User-Adapted Interaction*, *19*(1), 99-132. https://doi.org/10.1007/s11257-008-9047-z

[9] Clark, E. & Paran, A. (2007). The employability of non-native-speaker teachers of EFL: A UK survey. *System*, *35*(4), 407-430. https://doi.org/10.1016/j.system.2007.05.002

[10] Li, T. (2021). Application of APRIORI correlation algorithm on music education curriculum association rules. *Journal of Physics: Conference Series.* https://doi.org/10.1088/1742-6596/1955/1/012067

[11] Yuan, X. (2017). An improved Apriori algorithm for mining association rules. *AIP conference proceedings.* https://doi.org/10.1063/1.4977361

[12] Hegland, M. (2007). The apriori algorithm - a tutorial. *Mathematics and computation in imaging science and information processing*, 209-262. https://doi.org/10.1142/9789812709066_0006

[13] Pan, T. (2021). An Improved Apriori Algorithm for Association Mining Between Physical Fitness Indices of College Students. *International Journal of Emerging Technologies in Learning (IJET)*, *16*(9), 235-246. https://doi.org/10.3991/ijet.v16i09.22747

[14] Ibrahim, C. A. M. & Gaber, C. E. M. (2016). Knowledge Management and itsInfluence on the Efficiency of Maritime Education and Training Institutes (A Case Study). *International Association of Maritime Universities 17th Annual General Assembly 2016*, 144-157.

[15] Senčila, V. & Kalvaitienė, G. (2019). Industry 4.0: Autonomous Shipping and New Challenges for Maritime *Education and Training. Transport Means 2019-Sustainability: Research and Solutions - Proceedings of the 23rd International Scientific Conference.*

[16] Ahvenjärvi, S. (2017). Unmanned Ships And The Maritime Education And Training. *Global perspectives in MET: Towards Sustainable, Green and Integrated Maritime Transport.*

[17] Gilmartin, T., Gal, D., & O'Connor, E. A. (2019*). VR training videos: Using immersive technologies to support experiential learning methods in maritime education. IAMUC 2019-20th Commemorative Annual General Assembly, AGA 2019 - Proceedings of the International Association of Maritime Universities Conference.*

[18] Alop, A. (2019). The Challenges of the Digital Technology Era for Maritime Education and Training. *2019 European Navigation Conference (ENC).* https://doi.org/10.1109/EURONAV.2019.8714176

[19] Psaraftis, H. N. & Kontovas, C. A. (2010). Balancing the economic and environmental performance of maritime transportation. *Transportation Research Part D: Transport and Environment*, *15*(8), 458-462. https://doi.org/10.1016/j.trd.2010.05.001

[20] Linder, A. J. (2010). CO2 Restrictions and Cargo Throughput Limitations at California Ports: A Closer Look at AB 32 and Port-to-Port Shipping. *Public Works Management & Policy*, *14*(4), 374-391. https://doi.org/10.1177/1087724x10363811

[21] Chang, C. C. & Jhang, C. W. (2016). Reducing speed and fuel transfer of the Green Flag Incentive Program in Kaohsiung Port Taiwan. *Transportation Research Part D: Transport and Environment*, *46*, 1-10. https://doi.org/10.1016/j.trd.2016.03.007

[22] Woo, J.-K. & Moon, D. S.-H. (2013). The effects of slow steaming on the environmental performance in liner shipping. *Maritime Policy & Management*, *41*(2), 176-191. https://doi.org/10.1080/03088839.2013.819131

[23] ITF. (2019). *Container Shipping in Europe Data for the Evaluation of the EU Consortia Block Exemption*. Paris, France.

[24] Vasheghani, M. & Abtahi, M. (2022). Strategic planning for multimodal transportation in ports. *Maritime Policy & Management*, 1-23. https://doi.org/10.1080/03088839.2022.2061060

[25] ITF. (2019). *Maritime Subsidies: Do they Provide Value for Money?*

[26] Sebastian, S. (2016). Performance evaluation by artificial neural network using WEKA. *International Research Journal of Engineering and Technology*, *3*(3), 1459-1464.

[27] Borkar, S. & Rajeswari, K. (2013). Predicting students academic performance using education data mining. *International Journal of Computer Science and Mobile Computing*, *2*(7), 273-279.

[28] Alangari, N. & Alturki, R. (2020). Association rule mining in higher education: a case study of computer science students. *Smart Infrastructure and Applications*, 311-328. https://doi.org/10.1007/978-3-030-13705-2_13

[29] Hashima, A. S., Hamoud, A. K., & Awadh, W. A. (2018). Analyzing students' answers using association rule mining based on feature selection. *Journal of Southwest Jiaotong University*, *53*(5).

[30] Jha, J. & Ragha, L. (2013). Educational data mining using improved apriori algorithm. *International Journal of Information and Computation Technology*, *3*(5), 411-418.

[31] Jovanoski, V. & Lavrač, N. (2001). Classification rule learning with APRIORI-C. *Portuguese conference on artificial intelligence.* https://doi.org/10.1007/3-540-45329-6_8

[32] Pramod, S. & Vyas, O. (2012). Data stream mining: A review on windowing approach. *Global Journal of Computer Science and Technology Software & Data Engineering*, *12*(11), 26-30.

[33] Lakshmi, K. S. & Vadivu, G. (2017). Extracting association rules from medical health records using multi-criteria decision analysis. *Procedia Computer Science*, *115*, 290-295. https://doi.org/10.1016/j.procs.2017.09.137

[34] Altay, E. V. & Alatas, B. (2021). Differential evolution and sine cosine algorithm based novel hybrid multi-objective approaches for numerical association rule mining. *Information Sciences*, *554*, 198-221. https://doi.org/10.1016/j.ins.2020.12.055

Contact information:

Fang PENG
School of Maritime Economics and Management,
Dalian Maritime University,
No. 1, Linghai Road, Dalian, Liaoning, China
E-mail: pengfang@dlmu.edu.cn

Yuhui SUN
(Corresponding author)
UniSA STEM, University of South Australia,
Mawson Lakes Blvd, Mawson Lakes SA, 5095
E-mail: Yuhui.sun@mymail.unisa.edu.au

Zigen CHEN
School of Maritime Economics and Management,
Dalian Maritime University,
No. 1, Linghai Road, Dalian, Liaoning, China
E-mail: chenzigen2014@dlmu.edu.cn

Jing GAO
UniSA STEM, University of South Australia,
Mawson Lakes Blvd, Mawson Lakes SA, 5095
E-mail: Jing.gao@unisa.edu.au