INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATION & MANAGEMENT

IJRCM

A Monthly Double-Blind Peer Reviewed (Refereed/Juried) Open Access International e-Journal - Included in the International Serial Directories
Indexed & Listed at:
Ulrich's Periodicals Directory ©, ProQuest, U.S.A., EBSCO Publishing, U.S.A., Cabell's Directories of Publishing Opportunities, U.S.A.
as well as in Open J-Gage, India [link of the same is duly available at Inflibnet of University Grants Commission (U.G.C.)]
Registered & Listed at: Index Copernicus Publishers Panel, Poland
Circulated all over the world & Google has verified that scholars of more than 1500 Cities in 141 countries/territories are visiting our journal on regular basis.
Ground Floor, Building No. 1041-C-1, Devi Bhawan Bazar, JAGADHRI – 135 003, Yamunanagar, Haryana, INDIA
www.ijrcm.org.in

CONTENTS

CALL FOR MANUSCRIPTS

We invite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the area of Computer, Business, Finance, Marketing, Human Resource Management, General Management, Banking, Insurance, Corporate Governance and emerging paradigms in allied subjects like Accounting Education; Accounting Information Systems; Accounting Theory & Practice; Auditing; Behavioral Accounting; Behavioral Economics; Corporate Finance; Cost Accounting; Econometrics; Economic Development; Economic History; Financial Institutions & Markets; Financial Services; Fiscal Policy; Government & Non Profit Accounting; Industrial Organization; International Economics & Trade; International Finance; Macro Economics; Micro Economics; Monetary Policy; Portfolio & Security Analysis; Public Policy Economics; Real Estate; Regional Economics; Tax Accounting; Advertising & Promotion Management; Business Education; Management Information Systems (MIS); Business Law, Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labor Relations & Human Resource Management; Marketing Research; Marketing Theory & Applications; Non-Profit Organizations; Office Administration/Management; Operations Research/Statistics; Organizational Behavior & Theory; Organizational Development; Production/Operations; Public Administration; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism, Hospitality & Leisure; Transportation/Physical Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Digital Logic; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Multimedia; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic and Web Design. The above mentioned tracks are only indicative, and not exhaustive.

Anybody can submit the soft copy of his/her manuscript **anytime** in M.S. Word format after preparing the same as per our submission guidelines duly available on our website under the heading guidelines for submission, at the email address: **infoijrcm@gmail.com**.

GUIDELINES FOR SUBMISSION OF MANUSCRIPT

1. **COVERING LETTER FOR SUBMISSION:**

DATED: _____________

THE EDITOR
IJRCM

Subject: **SUBMISSION OF MANUSCRIPT IN THE AREA OF .**

(e.g. Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)

DEAR SIR/MADAM

Please find my submission of manuscript entitled '____' for possible publication in your journals.

I hereby affirm that the contents of this manuscript are original. Furthermore, it has neither been published elsewhere in any language fully or partly, nor is it under review for publication elsewhere.

I affirm that all the author (s) have seen and agreed to the submitted version of the manuscript and their inclusion of name (s) as co-author (s).

Also, if my/our manuscript is accepted, I/We agree to comply with the formalities as given on the website of the journal & you are free to publish our contribution in any of your journals.

NAME OF CORRESPONDING AUTHOR:
Designation:
Affiliation with full address, contact numbers & Pin Code:
Residential address with Pin Code:
Mobile Number (s):
Landline Number (s):
E-mail Address:
Alternate E-mail Address:

NOTES:
a) The whole manuscript is required to be in ***ONE MS WORD FILE*** only (pdf. version is liable to be rejected without any consideration), which will start from the covering letter, inside the manuscript.
b) The sender is required to mention the following in the **SUBJECT COLUMN** of the mail:
New Manuscript for Review in the area of (Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)
c) There is no need to give any text in the body of mail, except the cases where the author wishes to give any specific message w.r.t. to the manuscript.
d) The total size of the file containing the manuscript is required to be below **500 KB**.
e) Abstract alone will not be considered for review, and the author is required to submit the complete manuscript in the first instance.
f) The journal gives acknowledgement w.r.t. the receipt of every email and in case of non-receipt of acknowledgment from the journal, w.r.t. the submission of manuscript, within two days of submission, the corresponding author is required to demand for the same by sending separate mail to the journal.

2. **MANUSCRIPT TITLE**: The title of the paper should be in a 12 point Calibri Font. It should be bold typed, centered and fully capitalised.

3. **AUTHOR NAME (S) & AFFILIATIONS**: The author (s) **full name**, **designation**, **affiliation** (s), **address**, **mobile/landline numbers**, and **email/alternate email address** should be in italic & 11-point Calibri Font. It must be centered underneath the title.

4. **ABSTRACT**: Abstract should be in fully italicized text, not exceeding 250 words. The abstract must be informative and explain the background, aims, methods, results & conclusion in a single para. Abbreviations must be mentioned in full.

5. **KEYWORDS:** Abstract must be followed by a list of keywords, subject to the maximum of five. These should be arranged in alphabetic order separated by commas and full stops at the end.

6. **MANUSCRIPT**: Manuscript must be in ***BRITISH ENGLISH*** prepared on a standard A4 size ***PORTRAIT SETTING PAPER***. It must be prepared on a single space and single column with 1" margin set for top, bottom, left and right. It should be typed in 8 point Calibri Font with page numbers at the bottom and centre of every page. It should be free from grammatical, spelling and punctuation errors and must be thoroughly edited.

7. **HEADINGS**: All the headings should be in a 10 point Calibri Font. These must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.

8. **SUB-HEADINGS**: All the sub-headings should be in a 8 point Calibri Font. These must be bold-faced, aligned left and fully capitalised.

9. **MAIN TEXT**: The main text should follow the following sequence:

   INTRODUCTION

   REVIEW OF LITERATURE

   NEED/IMPORTANCE OF THE STUDY

   STATEMENT OF THE PROBLEM

   OBJECTIVES

   HYPOTHESES

   RESEARCH METHODOLOGY

   RESULTS & DISCUSSION

   FINDINGS

   RECOMMENDATIONS/SUGGESTIONS

   CONCLUSIONS

   SCOPE FOR FURTHER RESEARCH

   ACKNOWLEDGMENTS

   REFERENCES

   APPENDIX/ANNEXURE

   It should be in a 8 point Calibri Font, single spaced and justified. The manuscript should preferably not exceed ***5000 WORDS***.

10. **FIGURES & TABLES**: These should be simple, crystal clear, centered, separately numbered &self explained, and **titles must be above the table/figure**. **Sources of data should be mentioned below the table/figure**. It should be ensured that the tables/figures are referred to from the main text.

11. **EQUATIONS**:These should be consecutively numbered in parentheses, horizontally centered with equation number placed at the right.

12. **REFERENCES**: The list of all references should be alphabetically arranged. The author (s) should mention only the actually utilised references in the preparation of manuscript and they are supposed to follow **Harvard Style of Referencing**. The author (s) are supposed to follow the references as per the following:

- All works cited in the text (including sources for tables and figures) should be listed alphabetically.
- Use (**ed.**) for one editor, and (**ed.s)** for multiple editors.
- When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc, in chronologically ascending order.
- Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
- The title of books and journals should be in italics. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
- For titles in a language other than English, provide an English translation in parentheses.
- The location of endnotes within the text should be indicated by superscript numbers.

PLEASE USE THE FOLLOWING FOR STYLE AND PUNCTUATION IN REFERENCES:

BOOKS

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill, New Delhi.
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio" Ohio State University, Nigeria.

CONTRIBUTIONS TO BOOKS

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

JOURNAL AND OTHER ARTICLES

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.

CONFERENCE PAPERS

- Garg, Sambhav (2011): "Business Ethics" Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19–22 June.

UNPUBLISHED DISSERTATIONS AND THESES

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, KurukshetraUniversity, Kurukshetra.

ONLINE RESOURCES

- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

WEBSITES

- Garg, Bhavet (2011): Towards a New Natural Gas Policy, Political Weekly, Viewed on January 01, 2012 http://epw.in/user/viewabstract.jsp

A CONCEPT BASED APPROACH OF RARE ASSOCIATION RULE MINING FROM EDUCATION DATA

ASTHA PAREEK
RESEARCH SCHOLAR
DEPARTMENT OF COMPUTER SCIENCE & IT
THE IIS UNIVERSITY
JAIPUR

DR. MANISH GUPTA
DY. DIRECTOR (IT)
RCEE
DEPARTMENT OF EDUCATION
GOVERNMENT OF RAJASTHAN
JAIPUR

ABSTRACT
Data mining is the process of discovering useful knowledge in the form of patterns from the data. Association rule mining is an important knowledge discovery technique in the field of data mining. It involves finding interesting associations between the sets of objects in a transactional database. A rare association rule is an association rule with items having low support. In many real-world applications, rare association rules can provide useful information to the users. Rare association rules are those that only appear infrequently even though they are highly associated with very specific data. In consequence, these rules can be very appropriate for using with educational datasets since they are usually imbalanced. In this paper, we provide the basic concepts about rare association rule mining and survey the list of existing rare association rule mining techniques.

KEYWORDS
data mining, knowledge discover, association rule, rare association rule.

INTRODUCTION
Data mining (also known as Knowledge Discovery in Databases) has been defined as "the nontrivial extraction of implicit, previously unknown, and potentially useful information from data". Data mining uses machine learning, statistical and visualization techniques to discover and represent the knowledge in a form which is easily comprehensible to humans. Currently, data mining is being used in a wide range of industry applications, such as marketing, surveillance, fraud detection, and scientific discovery. The process of data mining mainly involves extraction of association rules, finding clusters, and assigning categories to data (classification). Currently, the active work in data mining has focused on discovering knowledge patterns from data streams time-series data, probabilistic databases and high-dimensional data .In addition, extraction of rare knowledge patterns (i.e., patterns involving rarely occurring entities) from different datasets is also an active research area in data mining .The main goal of association rule mining is to discover relationships among sets of items in a transactional database. Association rules aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in transaction databases or other data repositories. Association rules are if/then statements that help uncover relationships between seemingly unrelated data in a relational database or other information repository. An example of an association rule would be "If a customer buys a dozen eggs, he is 80% likely to also purchase milk." An association rule has two parts, an antecedent (if) and a consequent (then). An antecedent is an item found in the data. A consequent is an item that is found in combination with the antecedent. Association rules are created by analyzing data for frequent if/then patterns and using the criteria support and confidence to identify the most important relationships. Support is an indication of how frequently the items appear in the database. Confidence indicates the number of times the if/then statements have been found to be true. In data mining, association rules are useful for analyzing and predicting customer behavior. They play an important part in shopping basket data analysis, product clustering, and catalog design and store layout. Programmers use association rules to build programs capable of machine learning. Machine learning is a type of artificial intelligence (AI) that seeks to build programs with the ability to become more efficient without being explicitly programmed. A rare association rule is an association rule with items having low support. In many real-world applications, rare association rules can provide useful information to the users. Rare association rule mining has been hardly applied to educational data, despite the fact that infrequent associations can be of great interest since they are related to rare but crucial cases. For instance, they might allow the instructor to verify a set of rules concerning certain infrequent/abnormal learning problems that should be taken into account when dealing with students with special needs. Thus, this information could help the instructor to discover a minority of students who may need specific support with their learning process. From the perspective of knowledge discovery, the greatest reason for applying RARM in the field of education is the imbalanced nature of data in education, as in other real-world tasks, i.e., some classes have many more instances than others. Furthermore, in applications like education, the minor parts of an attribute can be more interesting than the major parts; for example, students who fail or drop out are usually less frequent than those students who fare well. In the field of association rule mining, the rare item problem is essentially considered to be a data imbalance problem which may, on either side of the association rule, give rise to severe problems. The problem of imbalance has only been dealt with in one educational data mining study. However, in this work, data was firstly modified/preprocessed to solve the problem of imbalance and then several different classification algorithms were applied instead of specific association rule algorithms.

CLASSIFICATION OF RARE ASSOCIATION
The rare associations are classified into two types.
1. Interesting rare association
2. Uninteresting rare association

INTERESTING RARE ASSOCIATION
An association is said to be interesting rare association if it has low support but the items contained in it are highly correlated or the confidence of the association is high.

UNINTERESTING RARE ASSOCIATION
An association is said to be uninteresting rare association if it has low support and low confidence. Usually the association rule mining algorithm consists of two steps: Finding the frequent Item sets and extracting the interesting association rules. This paper makes a combined effort to extract the interesting rare association rules and interesting frequent association rules in a single step.

REVIEW OF LITERATURE

Rare Association Rule Mining Detecting sporadic association rules, rules with low support but high confidence efficiently are a difficult data mining problem. To find these rules in traditional approaches, such as the Apriori algorithm, minimum support (minsup) has to be set very low, which results in a large amount of redundant rules. As a specific example of the problem, consider the association mining problem where we want to determine if there is an association between buying a food processor and buying a cooking pan (Liu et al. 1999).The problem is that both items are rarely purchased in a supermarket. Thus, even if the two items are almost always purchased together when either one is purchased, this association may not be found. Modifying the minsup threshold to take into account the importance of the items is one way to ensure that rare items remain in consideration. To find this association minsup must be set low. However setting this threshold low would cause a combinatorial explosion in the number of item sets generated. Frequently occurring items will be associated with one another in an enormous number of ways simply because the items are so common that they cannot help but appear together. This is known as the rare item problem. It means that using the Apriori algorithm, we are unlikely to generate rules that may indicate rare events of potentially dramatic Consequence. Some individual items can have such low support that they cannot contribute to rules generated by Apriori, even though they may participate in rules that have very high confidence. They overcome this problem with a technique called MSApriori whereby each item in the database can have a minimum item support (MIS) given byte user. By providing a different MIS for different items, a higher minimum support is tolerated for rules that involve frequent items and a lower minimum support for rules that involve less frequent items. Yun et al. (2003) proposed the RSAA algorithm to generate rules in which significant rare item sets take part, without any ”magic numbers” specified by the user. This technique uses relative support: RSup is used in place of support. Thus, this algorithm decreases the support threshold for items that have low frequency and increases the support threshold for items that have high frequency. Koh et al. (2008) proposed an approach to find rare rules with candidate item sets that fall below a maxsup (maximum support) level but above a minimum absolute support value. They introduced an algorithm called Apriori-Inverse to find sporadic rules efficiently: for example, a rare association of two common symptoms indicating a rare disease. They later proposed another approach called MIISR. In their approach, the consequent of these rules is an item below maxsup threshold and the antecedent has support below maxsup but may consist of individual items above maxsup. In both approaches they use minimum absolute support (minabssup) threshold value derived from an inverted Fisher's exact test to prune out noise. At the low levels of co-occurrences of candidate item sets that need to be evaluated to generate rare rules. The Fisher test provided a statistically rigorous method of evaluating significance of co-occurrences and was thus an integral part of their approach. Like Apriori and MSA priori, RSAA is exhaustive in its generation of rules, so it spends time looking for rules which are not sporadic (i.e. rules with high support and high confidence). If the minimum- allowable relative support value is set close to zero, RSAA takes a similar amount of time to that taken by Apriori generate low-support rules in amongst the high support rules.

MOTIVATION

After discovering frequent patterns, approaches based on“multiple minsups framework” use mincon f-based rule discovery technique proposed in [1] for mining association rules containing both frequent and rare items. However, mincon f constraint may not disclose truly interesting association rules [9, 10].

Example: Consider the following market-basketdata T from the grocery store, focusing on the purchase of tea and co f f ee. Let f(tea;co f f ee) = 20,f(tea) = 25, f(co f f ee) = 90 and |T| = 100. Using this data, we evaluate the association rule {tea} ⇒ {co f f ee} to have support=20% and confidence=80%. In other words, it can be said that out of all the people who drink tea, 80% of them drink coffee. However, 90% of all the people drink coffee regardless of the fact that they drink tea or not. Thus, the knowledge that one drinks tea decreases the chances of a customer drinking coffee from 90% to 80%. Thus the rule, {tea} ⇒ {co f f ee} is slightly misleading. As a result various interestingness measures, such as lift, correlation and all-confidence have been proposed for discovering useful association rules. Each measure has its own selection bias that justifies the rationale for preferring a set of association rules over another. As a result, selecting a right interestingness measure for mining association rules is a tricky problem. To confront this problem, a framework has been suggested in [11] for selecting a right measure. Inthis framework, authors have discussed various properties of a measure and suggested to choose a measure depending on the properties interesting to the user. In this paper, we make an effort to identify a set of properties that a user should consider for mining rare association rules.

CONCLUSION

Association rules containing rare items can provide useful knowledge to the user(s). Efforts are being made in the literature to mine frequent patterns containing both frequent and rare items. However, selecting a measure to generate rare association rules from the set of frequent patterns is an issue. This study presented an approach for finding both frequent and rare item set mining based on the Apriori framework. It uses automated item wise support thresholds for mining. This leaves the user free from finding an appropriate threshold for each dataset to run. These thresholds are automatically calculated and used by the algorithm.

REFERENCES

1. N. Jiang and L. Gruenwald. Research issues in data stream association rule mining. SIGMOD Rec., 35:14–19, March 2006.
2. Newman, D., Hettich, S., Blake, C. & Merz, C. (1998), 'UCI repository of machine learning databases', http://www.ics.uci.edu/~mlearn/MLRepository.html.
3. P.-N. Tan, V. Kumar, and J. Srivastava. Selecting the right interestingness measure for associationpatterns.InKDD, pages 32–41, 2002.
4. Plasse, M., Niang, N., Saporta, G., Villeminot, A. &Leblond, L. (2007), 'Combined use of associationrules mining and clustering methods to find relevant links between binary rare attributes in a large data set', Computational Statistics & Data Analysis52(1), 596–613.
5. S. Brin, R. Motwani, and C. Silverstein. Beyond market baskets: Generalizing association rules to correlations. In SIGMOD Conference, pages 265–276, 1997
6. W. J. Frawley, G. Piatetsky-shapiro, and C. J. Matheus. Knowledge discovery in databases: an overview, 1992.
7. W.Wu and L. Gruenwald. Research issues in mining multiple data streams. In Proceedings of the First International Workshop on Novel Data Stream Pattern Mining Techniques, Stream KDD '10, pages 56–60, New York, NY, USA, 2010. ACM.
8. Wang, K., Xu, C. & Liu, B. (1999), Clustering transactions using large items, in 'CIKM '99: Proceedings of the Eighth International Conference on Information and Knowledge Management', ACM Press, New York, NY, USA, pp. 483–490.
9. X. Wu, C. Zhang, and S. Zhang. Efficient miningof both positive and negative association rules. ACM Trans. Inf. Syst., 22(3):381–405,2004.
10. Xu, J., Xiong, H., Sung, S. Y. & Kumar, V. (2003),A new clustering algorithm for transaction data via caucus, in 'Advances in Knowledge Discovery and Data Mining: 7th Pacific-Asia Conference, PAKDD 2003, Seoul, Korea, April 30 - May 2, 2003. Proceedings', pp.551–562.
11. Y.-H. Hu and Y.-L. Chen. Mining association rules with multiple minimum supports: a new mining algorithm and a support tuning mechanism. Decision Support Systems, 42(1):1–24, 2006.

REQUEST FOR FEEDBACK

Dear Readers

At the very outset, International Journal of Research in Computer Application and Management (IJRCM) acknowledges & appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to request you to supply your critical comments and suggestions about the material published in this issue as well as on the journal as a whole, on our E-mail **infoijrcm@gmail.com** for further improvements in the interest of research.

If you have any queries please feel free to contact us on our E-mail **infoijrcm@gmail.com**.

I am sure that your feedback and deliberations would make future issues better – a result of our joint effort.

Looking forward an appropriate consideration.

With sincere regards

Thanking you profoundly

Academically yours

Sd/-

Co-ordinator

ABOUT THE JOURNAL

In this age of Commerce, Economics, Computer, I.T. & Management and cut throat competition, a group of intellectuals felt the need to have some platform, where young and budding managers and academicians could express their views and discuss the problems among their peers. This journal was conceived with this noble intention in view. This journal has been introduced to give an opportunity for expressing refined and innovative ideas in this field. It is our humble endeavour to provide a springboard to the upcoming specialists and give a chance to know about the latest in the sphere of research and knowledge. We have taken a small step and we hope that with the active co-operation of like-minded scholars, we shall be able to serve the society with our humble efforts.

Our Other Journals