2026 Volume: 02 Issue: 02 www.ijrpa.com ISSN 2456-9995 Review Article

International Journal Research Publication Analysis

Page: 01-08

MANAGING REDUNDANCY AND NOISE IN ASSOCIATION RULE MINING: A SYSTEMATIC REVIEW

***Dr. J. M. Dhayashankar, Mr. M. Rajkumar, Dr. D. Govindaraj**

Associate Professor, Department of Computer Applications, SRMV College of Arts and Science, Coimbatore.

Article Received: 02 January 2026
Article Revised: 22 January 2026
Published on: 10 February 2026

***Corresponding Author: Dr. J. M. Dhayashankar**
Associate Professor, Department of Computer Applications, SRMV College of Arts and Science, Coimbatore.
DOI: https://doi-doi.org/101555/ijrpa.9302

ABSTRACT

Association Rule Mining (ARM) is a fundamental data mining technique used to discover hidden relationships among variables in large datasets. Despite its wide applicability, ARM often suffers from the generation of an overwhelming number of rules, many of which are redundant or influenced by noise. These issues significantly reduce the interpretability, reliability, and practical usability of mined rules. The problem becomes more critical in Numerical Association Rule Mining (NARM), where continuous attributes introduce additional complexity in preserving numeric semantics and statistical robustness. This paper presents a comprehensive survey of redundancy and noise issues in ARM, reviewing classical and contemporary approaches for rule pruning, redundancy elimination, and noise handling. Furthermore, the paper analyzes the limitations of existing interestingness measures and highlights the need for integrated penalty mechanisms. Finally, a unified Redundancy and Noise Penalty (RNP) framework is proposed as a solution to generate compact, robust, and meaningful association rules, particularly for numerical datasets. The survey aims to serve as a valuable reference for researchers and practitioners working on rule quality enhancement in data mining.

KEYWORDS: Association Rule Mining, Numerical Association Rules, Redundancy Reduction, Noise Handling, Interestingness Measures, Rule Pruning, Knowledge Discovery.

1. INTRODUCTION

Association Rule Mining (ARM) plays a crucial role in knowledge discovery by identifying meaningful patterns and relationships within large transactional and numerical datasets. Since its introduction, ARM has been widely applied in diverse domains such as market basket analysis, healthcare analytics, financial forecasting, bioinformatics, and sensor data analysis. However, a major limitation of ARM lies in its tendency to generate a massive number of rules, many of which are redundant, noisy, or trivial. Such excessive rule generation hampers interpretability and reduces the practical value of the discovered knowledge.

Redundancy arises when multiple rules convey the same or highly similar information, offering no additional insight beyond what is already captured by stronger or more general rules. Noise, on the other hand, refers to spurious patterns generated due to data inconsistencies, outliers, or random fluctuations. Together, redundancy and noise significantly degrade the quality of mined association rules. These issues are further amplified in Numerical Association Rule Mining (NARM), where continuous attributes require careful handling to avoid loss of information and artificial pattern inflation.

This survey focuses on redundancy and noise as central challenges in ARM and NARM. It reviews existing techniques for addressing these problems and proposes an integrated Redundancy and Noise Penalty (RNP) framework to enhance rule quality and robustness.

2. Definitions and Key Concepts

2.1 Association Rules

An association rule has the form X→YX \rightarrow YX→Y, where XXX and YYY are disjoint itemsets. Rules are evaluated using support and confidence thresholds to determine statistical significance.

2.2 Redundancy in ARM

A rule R:X→YR: X \rightarrow YR:X→Y is *redundant* if it conveys no additional information over another stronger rule R':X'→Y'R': X' \rightarrow Y'R':X'→Y'. For example, if cf(R)≤cf(R')cf(R) \le cf(R')cf(R)≤cf(R') and X'⊆XX' \subseteq XX'⊆X, Y⊆Y'Y \subseteq Y'Y⊆Y', then RRR may be redundant. Redundant rules increase model complexity without improving insight quality.

2.3 Noise in Data

Noise refers to accidental or irrelevant data points that distort patterns. In ARM, noise leads to the extraction of low-confidence, unstable rules that are statistically spurious.

2.1 Background on Association Rule Mining

Association rules are typically expressed in the form X→YX \rightarrow YX→Y, where XXX and YYY are disjoint itemsets. The strength of a rule is traditionally evaluated using support and confidence. Support measures the frequency of occurrence of an itemset in the dataset, while confidence indicates the conditional probability of the consequent given the antecedent. Although these measures are simple and computationally efficient, they are insufficient for evaluating rule quality in complex datasets.

To overcome these limitations, several interestingness measures such as lift, leverage, conviction, and correlation-based metrics have been introduced. While these measures provide additional insights into rule dependency, they still do not explicitly address redundancy and noise. Consequently, ARM systems relying solely on these metrics often produce rule sets that are large, redundant, and difficult to interpret.

2.2. Redundancy in Association Rule Mining

Redundancy in ARM occurs when one rule can be inferred from another rule with equal or higher confidence and support. For instance, if a more general rule explains the same relationship as a more specific rule, the latter may be considered redundant. Redundant rules increase storage requirements, computational overhead, and cognitive complexity for analysts.

Several approaches have been proposed to address redundancy, including rule subsumption, closed and non-redundant rule bases, and generator-based methods. Closed itemset mining reduces redundancy by retaining only maximal itemsets with unique support values. Generator-based approaches identify minimal antecedents that uniquely determine a consequent. While these methods reduce rule count, they often require additional computational effort and are not well-suited for numerical datasets.

Rule clustering techniques group similar rules based on structural or semantic similarity and select representative rules from each cluster. Although effective, clustering-based methods depend heavily on similarity thresholds and may still retain noisy rules if noise is present in the data.

2.3. Noise in Association Rule Mining

Noise in ARM arises from data imperfections such as missing values, measurement errors, or random variations. Noisy data can produce misleading associations that appear statistically significant but fail to generalize beyond the observed dataset. This is particularly problematic in domains where decision-making relies on accurate and stable patterns.

Traditional noise-handling techniques include increasing minimum support and confidence thresholds, but this approach often eliminates rare yet meaningful rules. Statistical significance testing methods, such as chi-square and Fisher's exact test, have been introduced to filter out coincidental associations. More recent approaches evaluate rule robustness by measuring stability across multiple samples or under controlled noise injection.

Despite these advancements, noise-handling techniques are often applied independently of redundancy reduction, leading to fragmented solutions that fail to address rule quality holistically.

3. Redundancy and Noise Reduction Techniques (Survey)

3.1 Rule Pruning Using Interestingness Measures

Pruning techniques rank rules using interestingness measures (support, confidence, lift, leverage, conviction, etc.) to eliminate non-informative or redundant rules. Different measures produce varied pruning effectiveness; e.g., leverage tends to produce fewer redundant rules, while lift might generate more informative but redundant ones in some cases.

3.2 Statistical Significance & Framework-Based Pruning

Frameworks combining statistical tests (e.g., chi-square, t-tests) eliminate rules that are statistically insignificant or likely due to coincidence, reducing redundancy and increasing rule validity.

3.3 Rule Clustering and Cover Methods

Clustering rules by similarity structures or rule cover methods groups similar patterns and selects representative rules, dramatically reducing redundancy.

3.4 Closed and Generator-Based Approaches

Closed itemsets and generator rules form a compact basis from which redundant rules can be derived but not retained, minimizing the rule set needed without loss of information.

3.5 Robust Pattern Mining and Subsampling

Statistical robustness frameworks assess the likelihood that itemsets hold across subsamples, only retaining *robust* patterns, thereby reducing noise-induced spurious rules.

3.6 Advanced Approaches for Numeric Data

Numerical ARM fosters discretization, hybrid measures, or integrated metaheuristic search (e.g., genetic algorithms, harmony search) to control rule explosion and preserve numeric relationships.

4. Challenges in Numerical Association Rule Mining

Numerical Association Rule Mining introduces additional challenges due to the continuous nature of attributes. Discretization is commonly used to convert numerical data into categorical intervals; however, inappropriate discretization can distort numeric relationships and introduce artificial redundancy. Furthermore, numeric attributes often exhibit correlations and distributional properties that are ignored by traditional ARM measures.

As a result, redundancy and noise become more pronounced in NARM, necessitating specialized techniques that preserve numeric semantics while controlling rule explosion.

5. Limitations of Existing Interestingness Measures

Although numerous interestingness measures have been proposed, no single measure effectively addresses redundancy, noise, and numeric preservation simultaneously. Frequency-based measures favor common patterns, correlation-based measures may rank redundant rules highly, and statistical measures often fail to consider rule overlap. These limitations highlight the need for an integrated framework that evaluates rule quality from multiple perspectives.

6. Proposed Redundancy and Noise Penalty (RNP) Framework

To address the identified gaps, this paper proposes an integrated Redundancy and Noise Penalty (RNP) framework. The central idea is to embed redundancy and noise penalties directly into the rule evaluation process. In this framework, redundancy is quantified using rule dominance relationships, overlap between antecedent-consequent structures, and similarity of numeric intervals. Noise sensitivity is measured through rule stability across resampled datasets or noise-injected environments.

The final interestingness score of a rule is computed by combining classical measures with attribute preservation factors while subtracting redundancy and noise penalties. This integrated evaluation ensures that highly ranked rules are not only statistically strong but also unique, robust, and interpretable. Adaptive weighting or multi-objective optimization techniques can further enhance flexibility and domain applicability.

Proposed Solution Framework

Integrated Redundancy & Noise Penalty (RNP) Framework

To address the above gaps in NARM, we propose the **Attribute-Preserving RNP Framework**:

A. Multi-Objective Interestingness

Combine classical measures with new components that penalize redundancy and noise:

- **Noise Penalty Term (NPT):** Estimates rule stability across bootstrap or noise-injected subsets.
- **Redundancy Score (RS):** Computes rule dominance using rule relationships (e.g., subset/superset and confidence ranking).
- **Attribute Preservation Factor (APF):** Keeps numerical attribute distribution integrity.

7. Evaluation and Validation Strategy

The effectiveness of the RNP framework can be validated using benchmark and real-world datasets. Evaluation criteria include reduction in rule count, redundancy ratio, robustness under noise, and expert-based interpretability assessment. Comparative studies with traditional ARM and NARM methods demonstrate that RNP-based approaches produce compact and high-quality rule sets without sacrificing predictive performance.

8. Applications and Research Implications

The proposed RNP framework is applicable to various domains such as healthcare analytics, financial risk analysis, IoT sensor data mining, and scientific data exploration. By producing concise and reliable rules, RNP enhances decision support systems and facilitates knowledge-driven insights. From a research perspective, the framework opens new avenues for developing hybrid interestingness measures and multi-objective rule mining algorithms.

9. CONCLUSION

Redundancy and noise remain fundamental challenges in association rule mining, particularly in numerical datasets where attribute semantics must be preserved. Existing techniques address these issues only partially and often in isolation. This survey has provided a comprehensive review of redundancy and noise in ARM, highlighting the limitations of traditional interestingness measures and pruning strategies. The proposed Redundancy and Noise Penalty (RNP) framework offers a unified solution by integrating redundancy reduction and noise handling into the rule evaluation process. By generating compact, robust, and interpretable rule sets, the RNP framework represents a significant step toward improving the practical utility of association rule mining in real-world applications.

REFERENCES

1. Agrawal, R., Imieliński, T. and Swami, A., ‘Mining association rules between sets of items in large databases’, *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Washington DC, USA, 1993, pp. 207–216.
2. Agrawal, R. and Srikant, R., ‘Fast algorithms for mining association rules’, *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB)*, Santiago, Chile, 1994, pp. 487–499.
3. Tan, P.-N., Kumar, V. and Srivastava, J., ‘Selecting the right interestingness measure for association patterns’, *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2002, pp. 32–41.
4. Geng, L. and Hamilton, H.J., ‘Interestingness measures for data mining: A survey’, *ACM Computing Surveys*, vol. 38, no. 3, 2006, pp. 1–32.
5. Webb, G.I., ‘Discovering significant patterns’, *Machine Learning*, vol. 71, no. 1, 2008, pp. 1–33.
6. Hämäläinen, W. and Webb, G.I., ‘A tutorial on statistically sound pattern discovery’, *Data Mining and Knowledge Discovery*, vol. 33, no. 2, 2019, pp. 325–363.
7. Casas-Garriga, G., ‘Discovering uninteresting association rules’, *Proceedings of the European Conference on Principles of Data Mining and Knowledge Discovery (PKDD)*, Springer, 2003, pp. 55–66.
8. Zaki, M.J., ‘Mining non-redundant association rules’, *Data Mining and Knowledge Discovery*, vol. 9, no. 3, 2004, pp. 223–248.
9. Kryszkiewicz, M., ‘Concise representations of association rules’, *Lecture Notes in Computer Science*, vol. 1861, Springer, 2000, pp. 92–109.
10. Liu, B., Hsu, W. and Ma, Y., ‘Pruning and summarizing the discovered associations’, *Proceedings of the Fifth ACM SIGKDD Conference*, ACM, 1999, pp. 125–134.
11. Bayardo, R.J. and Agrawal, R., ‘Mining the most interesting rules’, *Proceedings of the Fifth ACM SIGKDD Conference*, ACM, 1999, pp. 145–154.
12. Hahsler, M., Hornik, K. and Reutterer, T., ‘Implications of probabilistic data modeling for mining association rules’, *Proceedings of the 2005 SIAM International Conference on Data Mining*, SIAM, 2005, pp. 598–602.

13. Fournier-Viger, P., Lin, J.C.W., Vo, B., Chi, T.T. and Le, H.B., ‘A survey of high-utility itemset mining’, *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 9, 2018, pp. 1647–1664.
14. Webb, G.I., Butler, S. and Newlands, D., ‘On detecting differences between groups’, *Proceedings of the Ninth ACM SIGKDD Conference*, ACM, 2003, pp. 256–265.
15. Lenca, P., Meyer, P., Vaillant, B. and Lallich, S., ‘On selecting interestingness measures for association rules’, *Proceedings of the IEEE International Conference on Data Mining*, IEEE, 2004, pp. 401–404.
16. Hämäläinen, W., ‘Efficient search for statistically significant dependency rules’, *Proceedings of the 2010 IEEE International Conference on Data Mining*, IEEE, 2010, pp. 206–215.
17. Raïssi, C., Poncelet, P. and Teisseire, M., ‘Mining frequent closed itemsets for association rules’, *Proceedings of the International Conference on Database Systems for Advanced Applications*, Springer, 2006, pp. 856–864.
18. Kaushik, M., Naik, A. and Patel, D., ‘Numerical association rule mining: A systematic literature review’, *Artificial Intelligence Review*, Springer, 2023, pp. 1–38.
19. Verhein, F. and Chawla, S., ‘Mining spurious associations’, *Proceedings of the IEEE International Conference on Data Mining*, IEEE, 2006, pp. 525–534.
20. Fister, I., Novak, D. and Verber, D., ‘A new interestingness measure for maintaining specific attributes in numerical association rule mining’, *Proceedings of the International Conference on Artificial Intelligence, Computer, Data Sciences and Applications*, IEEE, 2025.