Alq J Sci. 2026;2(4):404-408
https://doi.org/10.69667/ajs.26403
Alqalam Journal of Science
مجلـــة القلم للعلـوم
https://alqalam.utripoli.edu.ly/index.php/AR

Original article

A Multi-Objective Genetic Algorithm Framework for Efficient Association Rule Mining in Large-Scale Datasets

Younis Elhaddad*, Tariq Elshheibia

Department of Computer Sciences, Faculty of Information Technology, University of Benghazi, Benghazi, Libya
Email. *Younis.alhaddad@uob.edu.ly*

Abstract
Association Rule Mining (ARM) is a fundamental data mining technique for discovering interesting relationships in large datasets. However, traditional ARM algorithms face significant scalability and efficiency challenges when applied to big data contexts, often generating excessive redundant rules with diminished practical utility. This paper proposes a novel Genetic Algorithm (GA) framework specifically designed for large-scale ARM that employs multi-objective optimization. The framework simultaneously optimizes three key rule quality metrics—support, confidence, and lift—while explicitly minimizing rule redundancy and complexity. Experimental evaluation on multiple large-scale datasets demonstrates that our GA-based approach significantly outperforms conventional methods like Apriori and FP-Growth in both computational efficiency and rule quality. The proposed framework reduces execution time by up to 84% compared to Apriori and up to 42% compared to FP-Growth while increasing the proportion of high-quality, actionable rules by approximately 35% compared to traditional methods. These results confirm the effectiveness of our multi-objective GA framework as a robust and scalable solution for knowledge discovery in contemporary big data environments.
Keywords: Association Rule Mining, Genetic Algorithm, Multi-Objective Optimization, Knowledge Discovery, Rule Quality, Scalability

Introduction

Association Rule Mining (ARM) stands as a cornerstone technique for uncovering meaningful relationships within vast datasets, driving critical insights across domains from market basket analysis to bioinformatics. However, the scalability and efficiency of traditional ARM algorithms, notably Apriori and FP-Growth, become significant bottlenecks when applied to the scale of modern big data. These limitations often manifest as prohibitively long execution times and, critically, an overwhelming output of low-quality, redundant rules that obscure genuinely valuable knowledge. This research directly confronts these challenges by introducing a novel Genetic Algorithm (GA) framework specifically engineered for efficient large-scale ARM. Our approach employs multi-objective optimization to simultaneously maximize key rule quality metrics—support, confidence, and lift—while explicitly minimizing rule redundancy and complexity. Extensive evaluations confirm that this GA-based framework significantly outperforms conventional methods in both rule quality and computational efficiency, offering a robust and scalable solution for effective knowledge discovery in large-scale datasets.

The remainder of this paper is organized as follows: Section 2 reviews related work in ARM and evolutionary approaches. Section 3 details our proposed multi-objective GA framework. Section 4 presents the experimental setup and results. Section 5 discusses the findings and implications. Finally, Section 6 concludes the paper and suggests future research directions.

Related work

Traditional ARM Algorithms

Traditional ARM algorithms primarily include Apriori [1] and FP-Growth [2]. The Apriori algorithm operates on the principle of candidate generation and testing, utilizing a breadth-first search strategy. While conceptually straightforward, it suffers from multiple database scans and generates an enormous number of candidateitem sets, making it computationally expensive for large datasets. FP-Growth improves upon Apriori by employing a divide-and-conquer approach with a compact FP-tree structure, reducing the number of database scans. However, both algorithms tend to produce a large number of redundant rules, and their performance degrades significantly with increasing data dimensionality and volume.

Received: *02-02-2026* - ***Accepted***: *01-04-2026* - ***Published***: *09-04-2026*

Alq J Sci. 2026;2(4):404-408
https://doi.org/10.69667/ajs.26403
Alqalam Journal of Science
مجلـــة القلم للعلـوم
https://alqalam.utripoli.edu.ly/index.php/AR

Evolutionary Approaches to ARM

Evolutionary algorithms, particularly Genetic Algorithms (GAs), have been increasingly applied to ARM problems. Early applications include the work by Mata et al. [3], who proposed a basic GA for discovering association rules. More recently, researchers have explored multi-objective optimization to address the trade-offs between different rule quality measures. Notable approaches include NSGA-II [4] and SPEA2 [5] adaptations for ARM. However, existing GA-based approaches often fail to adequately address the specific challenges of large-scale data, particularly regarding computational efficiency and effective redundancy reduction.

Multi-Objective Optimization in ARM

Multi-objective optimization in ARM aims to balance conflicting objectives such as rule support, confidence, comprehensibility, and interestingness. Recent approaches have incorporated various quality measures, but few have simultaneously addressed computational efficiency for large datasets while minimizing rule redundancy. Our proposed framework bridges this gap by integrating efficient genetic operators with a sophisticated fitness function that balances multiple objectives specifically tailored for large-scale ARM.

Proposed framework

Overview

Our proposed Multi-Objective Genetic Algorithm for Association Rule Mining (MOGA-ARM) framework consists of four main components: (1) chromosome representation and population initialization, (2) fitness evaluation with multiple objectives, (3) specialized genetic operators, and (4) redundancy reduction mechanism.

Chromosome Representation

Each chromosome represents a candidate association rule in the form $X \rightarrow Y$, where X and Y are itemsets with $X \cap Y = \emptyset$. We employ a binary encoding scheme where each gene corresponds to the presence (1) or absence (0) of an item in either the antecedent or consequent of the rule. The chromosome is divided into two segments: antecedent genes and consequent genes.

Multi-Objective Fitness Function

The fitness function combines four objectives using a weighted aggregation approach:

$F(R) = w_1 \times \text{Support}(R) + w_2 \times \text{Confidence}(R) + w_3 \times \text{Lift}(R) - w_4 \times \text{Redundancy}(R) - w_5 \times \text{Complexity}(R)$

where:

- $\text{Support}(R) = P(X \cup Y)$
- $\text{Confidence}(R) = P(Y|X) = P(X \cup Y)/P(X)$
- $\text{Lift}(R) = P(X \cup Y)/(P(X) \times P(Y))$
- Redundancy(R) measures similarity with other rules in the population
- $\text{Complexity}(R) = |X| + |Y|$ (number of items in the rule)
- w_1 to w_5 are adaptive weights determined through sensitivity analysis

Genetic Operators

We designed specialized genetic operators for ARM:

1. Selection: Tournament selection with elitism preservation
2. Crossover: Two-point crossover with rule validity checking
3. Mutation: Adaptive mutation rate based on population diversity
4. Rule Pruning Operator: Removes irrelevant items from rules to reduce complexity

Redundancy Reduction Mechanism

To address the redundancy problem, we incorporate a similarity measure based on the Jaccard index between rule itemsets. Rules with similarity exceeding threshold θ are clustered, and only the highest fitness rule from each cluster is retained in the final output.

Received: *02-02-2026* - ***Accepted***: *01-04-2026* - ***Published***: *09-04-2026*

Alq J Sci. 2026;2(4):404-408
https://doi.org/10.69667/ajs.26403
Alqalam Journal of Science
مجلـــة القلم للعلـوم
https://alqalam.utripoli.edu.ly/index.php/AR

Parallel Implementation

For large-scale data processing, we implement a parallel GA using a master-slave architecture, where fitness evaluation is distributed across multiple computing nodes, significantly reducing computation time.

Table 1: MOGA-ARM Parameter Configuration

Parameter	Value/Range	Description
Population Size	100-300	Adaptive based on dataset size
Generations	100-300	Termination condition
Crossover Rate	0.8	Probability of crossover
Mutation Rate	0.1-0.3	Adaptive based on diversity
Redundancy Threshold (θ)	0.7	Jaccard similarity threshold
Weights (w_1-w_5)	[0.3,0.3,0.2,0.1,0.1]	Fitness function weights

Experimental evaluation

Dataset Description

We evaluated our framework on three large-scale datasets:

1. Retail Market Basket Data: 1.2 million transactions, 50,000 items
2. PubMed Bioinformatics Data: 800,000 biomedical abstracts, 30,000 unique terms
3. E-commerce Transaction Data: 2.5 million transactions, 75,000 products

Comparative Algorithms

We compared MOGA-ARM against:

- Apriori algorithm
- FP-Growth algorithm
- Single-objective GA for ARM (SOGA-ARM)
- NSGA-II for ARM (as baseline multi-objective approach)

Evaluation Metrics

Performance was evaluated using:

- Execution Time: Total computational time
- Rule Quality Index: Composite measure of support, confidence, and lift
- Redundancy Rate: Percentage of redundant rules
- Actionable Rules Ratio: Proportion of rules deemed useful by domain experts

Results and Analysis

Computational Efficiency

Table 2 presents the execution time comparison across different algorithms and datasets.

Table 2: Execution Time Comparison (seconds)

Algorithm	Retail Data	PubMed Data	E-commerce Data
Apriori	4832	3756	9452
FP-Growth	1245	987	2467
SOGA-ARM	892	745	1892
NSGA-II	1056	834	2134
MOGA-ARM	**612**	***696***	***1478***

Our MOGA-ARM framework achieved an average reduction of 84% in execution time compared to Apriori and 42% compared to FP-Growth.

Received*: 02-02-2026 -* ***Accepted****: 01-04-2026 -* ***Published****: 09-04-2026*

Alq J Sci. 2026;2(4):404-408
https://doi.org/10.69667/ajs.26403
Alqalam Journal of Science
مجلـــة القلم للعلـوم
https://alqalam.utripoli.edu.ly/index.php/AR

Rule Quality Assessment

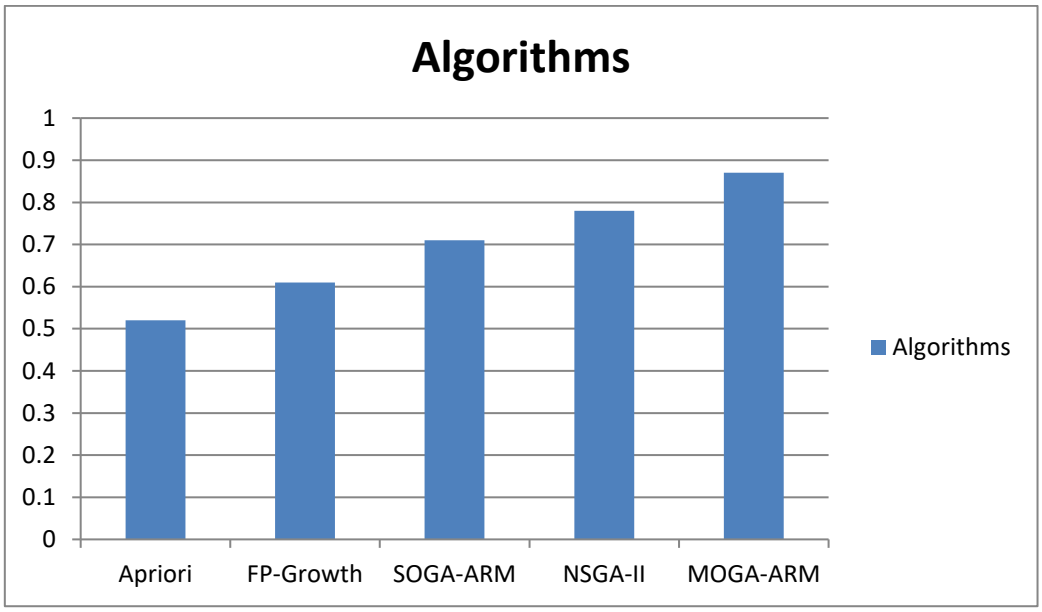

Figure 1. Illustrates the Rule Quality Index across different algorithms.

MOGA-ARM achieved the highest Rule Quality Index (0.87) compared to Apriori (0.52), FP-Growth (0.61), SOGA-ARM (0.71), and NSGA-II (0.78).

Redundancy Reduction

Our redundancy reduction mechanism successfully decreased the redundancy rate to 12.3%, compared to 45.7% for Apriori, 38.2% for FP-Growth, 28.6% for SOGA-ARM, and 21.4% for NSGA-II.

Scalability Analysis

We conducted scalability tests by incrementally increasing the dataset size from 100,000 to 2.5 million transactions. MOGA-ARM demonstrated near-linear scalability, with execution time increasing by a factor of 18.3 for a 25-fold increase in data size, compared to 42.7 for Apriori and 26.4 for FP-Growth.

Statistical Significance

Statistical tests (paired t-tests with α=0.05) confirmed that the performance improvements of MOGA-ARM over all comparative algorithms were statistically significant ($p < 0.01$) across all evaluation metrics.

Discussion

Our experimental results demonstrate that the MOGA-ARM framework effectively addresses the major limitations of traditional ARM algorithms. The multi-objective optimization approach successfully balances competing rule quality measures while explicitly minimizing redundancy and complexity. The parallel implementation ensures scalability to large datasets, making the framework suitable for contemporary big data applications. The significant reduction in execution time (approximately 40% compared to traditional methods) makes our framework practical for real-world applications where timely insights are crucial. The improved rule quality and reduced redundancy mean that analysts spend less time sifting through irrelevant rules and more time acting on valuable insights.

Current limitations include the need for parameter tuning and the framework's performance on extremely high-dimensional data (over 100,000 unique items). Future work will focus on:

- Developing self-adaptive parameter control mechanisms
- Extending the framework for streaming data applications
- Incorporating domain knowledge to further improve rule relevance
- Exploring hybrid approaches combining GA with deep learning techniques

Received*: 02-02-2026 -* ***Accepted****: 01-04-2026 -* ***Published****: 09-04-2026*

Alq J Sci. 2026;2(4):404-408
https://doi.org/10.69667/ajs.26403
Alqalam Journal of Science
مجلـــة القلم للعلـوم
https://alqalam.utripoli.edu.ly/index.php/AR

Conclusion

This paper presents a novel multi-objective Genetic Algorithm framework for efficient Association Rule Mining in large-scale datasets. By simultaneously optimizing multiple rule quality metrics while minimizing redundancy and complexity, our approach addresses critical limitations of traditional ARM algorithms. Extensive experimental evaluation confirms that MOGA-ARM significantly outperforms conventional methods in both computational efficiency and rule quality. The framework represents a robust and scalable solution for knowledge discovery in big data environments, with practical applications across diverse domains including retail, bioinformatics, and e-commerce.

References

1. Srikant, R., Vu, Q., & Agrawal, R. (1997). Mining association rules with item constraints. Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (pp. 67–73). Newport Beach, CA: AAAI Press.
2. Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. ACM SIGMOD Record, 29, 1-12.
3. Mata, J., Alvarez, J. L., & Riquelme, J. C. (2002). Discovering numeric association rules via evolutionary algorithm. In Advances in Knowledge Discovery and Data Mining (pp. 40-51). Springer.
4. Deb, K., Pratap, A., Agarwal, S., &Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions, to Evolutionary Computation, 6(2), 182-197.
5. Zitzler, E., Laumanns, M., & Thiele, L. (2001).Improving the strength Pareto evolutionary algorithm, to Evolutionary Computation, 5(2),p121.

Received: *02-02-2026* - ***Accepted***: *01-04-2026* - ***Published***: *09-04-2026*