Soft Set ARM Yields Compact Interpretable Patterns Under Uncertainty

Zahra Arwananing Tyas[1*], Mustafa Mat Deris[1], Iwan Tri Riyadi Yanto[2]

[1]Department of Information Technology, Universiti Muhammadiyah Malaysia, Perlis, Malaysia
[2]Department of Information Systems, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

[*]Corresponding Author Email: p5240011@student.umam.edu.my

Abstract. General Background Association rule mining is widely used to identify co-occurrence structures in transactional and event data, but classical Boolean formulations are limited when data are uncertain, incomplete, multi-valued, hierarchical, or temporally varying. Specific Background Soft set theory offers a parameterized representation of uncertainty and has been integrated with association rule mining to address these limitations without relying on membership functions or equivalence relations. Knowledge Gap However, the literature remains fragmented across foundational models, algorithmic variants, hybrid formulations, and application domains, making the overall development of this area difficult to assess systematically. Aims This paper reviews and synthesizes 24 Scopus-indexed studies to clarify conceptual foundations, organize integration strategies, map application areas, and identify open research gaps. Results The synthesis shows that soft-set-based association rule mining can reproduce or improve classical and rough approaches, reduce runtime, and control rule proliferation through maximal rules and parameter reduction. Hybrid fuzzy-soft, temporal, and N-soft formulations also support real-valued, noisy, time-aware, and multi-group data across education, healthcare, logistics, bioinformatics, and web and text analytics. Novelty The study provides an integrated synthesis across foundational formulations, algorithmic enhancements, hybrid models, and domain-specific applications within a single review framework. Implications These findings position soft-set-based association rule mining as a mature and versatile framework for decision support while highlighting the need for scalable algorithms, automatic parameter learning, deeper hybridization, and rigorous benchmarking.

Keywords: Soft Sets, Association Rules, Maximal Rules

1 Introduction

The rapid growth of digital traces in e-commerce, healthcare, finance, and social platforms has made automated knowledge discovery a core capability of modern information systems [1], [2], [3], [4]. Association rules mining (ARM) is one of the most widely used tools for uncovering co-occurrence patterns in such data, typically

R. Lomotey et al. (eds.), *Proceedings of the 1st International Conference on Communication and Digital Multimedia 2025 (ICCDM 2025)*, Advances in Social Science, Education and Humanities Research 1020,
https://doi.org/10.2991/978-2-38476-589-8_61

built on Boolean transaction models where an item either appears or does not appear in a transaction and attributes either satisfy or violate a predicate [5]. This crisp representation is effective for many transactional databases, but it becomes increasingly restrictive when data are uncertain, incomplete, multi-valued, or parameter-dependent—for example, when attributes are measured on quantitative scales, when descriptors are linguistic, or when the relevance of an attribute depends on a specific context or time period [6], [7], [8]. Comparative studies of nondeterministic set models and big-data–oriented ARM frameworks highlight that classical crisp rules struggle to represent vagueness, overlapping categories, and context-sensitive patterns that are common in modern applications [9], [10].

To address uncertainty and ambiguity, several mathematical frameworks, most prominently fuzzy sets, rough sets, and soft sets, have been integrated into ARM. Fuzzy and rough sets have been successfully applied for many years, especially for classification, clustering, feature reduction, and association rule discovery in noisy or imprecise environments [11], [12], [13]. However, both require additional structures such as membership functions or equivalence relations, which can be difficult to specify and tune for high-dimensional, heterogeneous data [11], [14]. Soft set theory, introduced by Molodtsov as a parameterized family of subsets, offers an alternative in which uncertainty is modeled through parameters rather than membership grades, and is “free from the difficulties affecting the existing methods” of probability, fuzzy, and rough sets [15]. In data-mining contexts, soft sets are particularly attractive because they can encode multi-valued, context-aware, and hierarchically structured attributes through flexible parameterizations and logical formulas defined over soft approximations [15], [16], [17].

To address ambiguity and contextual variation, ARM has increasingly been coupled with mathematical frameworks that model uncertainty. Among these, soft set theory offers a parameterization-based representation that can encode multi-valued, context-aware, and hierarchically structured attributes without committing to graded memberships. Evidence from the literature shows that transforming transactions into soft sets and mining under soft semantics can (i) reproduce the maximal rules obtained by traditional and rough-set approaches while (ii) substantially reducing execution time; subsequent studies refine efficiency via dedicated tree structures and pruning criteria, and extend the paradigm to temporal, hierarchical, and hybrid fuzzy–soft settings—spanning applications from text and air quality to logistics, education, bioinformatics, and clinical decision support.

This review uses 24 papers form Scopus with keyword “soft set” AND “association rules”, has goals (i) Conceptual grounding that can clarify soft set fundamentals and relate them to ARM primitives (ii) description that systematize how soft sets plug into ARM: parameter/attribute reduction for dimensionality control; soft definitions of support–confidence; multi-soft encodings aligned with multi-level ARM; temporal soft sets for time-scoped rules; constraint-based and maximal mining; and hybrids with fuzzy/rough models for real-valued or noisy data. Empirical synthesis results in across domains and identify open problems: scalable distributed implementations for large multi-soft encodings, standardized evaluation protocols balancing accuracy–redundancy–interpretability, and underexplored metaheuristic pipelines (e.g., GA-optimized soft-set ARM). Paper organization. Section 2 reviews soft set fundamentals; Section 3 presents integration patterns with ARM; Section 4 surveys applications and comparative findings; Section 5 concludes with practical recommendations for deploying soft-set-enhanced ARM.

2 Fundamental Concept of Soft Set Theory and Association Rules

In this section, we describe the concept and notation soft set of Molodtsov and association rule of Agrawal.

2.1 Soft Set Theory

A Soft set defined by F: E P(U) where U is the initial universe, P(U) is power set of U, E is the set of parameters. Let $S = (U, A, V, f)$ be a categorical-valued information system, where $U = \{u_1, u_2, \dots, u_n\}$ is a finite set of instance, $A = \{a_1, a_2, \dots, a_m\}$ is a finite set of the attribute, is values set of each attribute A, f is mapping function $f: (U, A) \rightarrow V$ and $S = (U, a_i, V_{a_i}, f)$, i = 1,2,…,|A| is Boolean-valued information system. [18], [19], [20]

2.2 Association Rules

Following the research of [21] ARM are expressed $X \Longrightarrow Y$, where $X, Y \subseteq I$ and $X \cap Y = \emptyset$ and I is the set of items in the transaction database D. In ARM has two fundamental metrics are support and confidence to enhance the evaluation process [22]. Support determines the frequency of appearance of a particular collection of items in a given dataset. This specific metric is essential for understanding the pervasiveness of a set of items or rules defined in a dataset. In other words, support indicates the proportion of transactions in the dataset that include a specific item configuration (the antecedents and consequences of a rule). Increased support indicates that this collection of items is common in the dataset.

$$Support\ (X) = \frac{Number\ of\ transactions\ containing\ itemset\ X}{Total\ number\ of\ transactions} \tag{1}$$

Equation 1 is support formula where X can represent individual items or groups. An increase in support indicates that item appears frequently in the dataset. A decrease in support may indicate that the rule is too specific or not general enough to be useful in practice.

Confidence evaluates the dependability of association rules, equation 2 is confidence formula. This metric quantifies the conditional probability that the consequent of a rule (right side) will appear, given the confirmed presence of the antecedent (left side). It addresses the probability of Y's occurrence under the condition that X has already occurred.

$$Confidence\ (X\ \rightarrow Y) = \frac{Support\ (X \cup Y)}{Support\ (X)} \quad (2)$$

Where:
X : antecedent (left side of the rule)
Y : consequent (right side of the rule)
Support(X∪Y) : the support of the itemset containing both X and Y

3 Integration and Application

Over roughly the last decade and a half, a substantial body of work has integrated soft set theory with ARM and its variants in order to better handle vague and complex data [1], [2], [3], [4], [5], [6], [7], [9], [10], [18], [19], [23], [24], [25], [26]. Foundational contributions introduced soft-set–based maximal association rules and general soft-set ARM frameworks and showed that soft maximal rules can match or improve the quality of rules obtained via rough sets while reducing search space and runtime [23], [24], [25]. Subsequent studies extended these ideas in several directions: maximal association analysis via logical formulas over soft sets, soft maximal rules for web usage mining, and maximal soft rules for text data demonstrate how parameterized soft descriptions support efficient rule mining in large transactional, web, and document collections [7], [16], [18]. Fuzzy soft and soft–fuzzy set models further enrich the representation by combining soft parameterization with fuzzy membership, allowing association rules to be extracted from real-valued or graded attributes, such as quantitative features in text or biological sequences [6], [19], [26], [27]. Empirical applications span student skill profiling, text classification, web user behavior, peptide sequence analysis for Mycobacterium tuberculosis and dengue virus, early warning of maritime logistics risks, temporal pattern discovery, promotion strategies in distance education, and heart-disease risk modeling, illustrating that soft-set-based ARM can capture subtle, domain-specific associations that classical crisp rules often miss [1], [2], [3], [4], [7], [16], [18], [19], [26], [27].

Taken together, these studies indicate that soft set theory provides a versatile and powerful extension of classical ARM, capable of modeling multi-valued, context-dependent, and temporally enriched attributes while remaining conceptually simple and computationally tractable [2], [7], [8], [15], [16], [17]. At the same time, existing contributions are fragmented across different variants (fuzzy soft sets, soft–fuzzy sets, temporal soft sets, N-soft sets, pliancy tree soft sets) and application domains, and several papers explicitly note that soft-set-based ARM still "needs more attention from both theoretical and practical side," including hybridization with metaheuristics and validation on large real-world data [2], [10], [15], [23], [24]. Therefore, research on soft set theory for association rules is not only active but also highly promising for further development. This paper responds to that need by systematically reviewing soft-set-based approaches to association rules, organizing them along three axes—conceptual foundations, integration patterns with other mathematical frameworks and algorithms, and application domains—and by identifying open challenges and future research directions that can guide the next generation of soft-set-driven association rule mining methods.

Based on this body of work, several concrete real case studies emerge where soft-set-based association rule mining can be directly applied or extended: (1) adaptive promotion and retention strategies in higher education, modeling institutional data with N-soft sets and mining rules over learner demographics, engagement, and support services [1], [9]; (2) clinical decision support for chronic diseases, generalizing pliancy tree soft sets from heart disease to other conditions (e.g., diabetes, hypertension) to capture complex interactions among lab values, lifestyle factors, and comorbidities [4]; (3) early-warning systems in logistics and supply chains, integrating qualitative risk assessments with quantitative indicators using soft-set association rules, as in maritime logistics but extended to other transport modes [3]; (4) temporal behavior analysis in retail or sensor networks, using temporal soft sets to discover time-aware associations between promotions and purchases or between sensor anomalies and equipment failures [5]; (5) student competency mapping and curriculum analytics, where fuzzy-soft association rules over student attributes and outcomes inform curriculum redesign and remedial support [1], [2], [9], [10]; (6) bioinformatics and pathogen characterization, extending fuzzy-soft and soft-fuzzy amino-acid association mining from MTBC and dengue to other pathogens and protein families [11], [15], [23], [25]; and (7) text mining and web usage mining, where maximal soft-set-based association rules—validated on text corpora and air-pollution data—are used to analyze keyword co-occurrence, search logs, and navigation paths on large websites [6], [7], [14], [17], [18], [19], [24].

Overall, the integration of soft set theory and association rule mining has produced a coherent methodological trajectory: (i) foundational formulations that map transactional data to soft sets and define soft association and maximal rules [18], [19], [27]; (ii) algorithmic enhancements for efficiency and pruning, including specialized trees, CRS, and clustering-assisted or constraint-based soft sets [6], [7], [13], [14], [17]; (iii) hybrid models combining soft sets with fuzzy sets, rough sets, and, prospectively, genetic algorithms [1], [2], [4], [10], [11], [15], [16], [23], [25], [26], [27]; (iv) successful applications in domains where data are incomplete, multi-valued, or semantically rich [1], [2], [3], [4], [9], [10], [11], [15], [18], [19], [25], [26], [27]; The literature also reveals an opportunity: while soft sets have been effectively combined with fuzzy and rough sets and optimization ideas have been outlined, full-fledged, large-scale hybrid systems—for example, GA-optimized fuzzy-soft association rules over temporal N-soft sets—are still rare [6], [7], [16], [26], [27]. This gap provides a clear direction for future work, especially for designing interpretable, domain-adapted rule-based systems in education, healthcare, logistics, and other high-uncertainty environments.

4 Conclusion and Future Research

Across the 24 reviewed papers, soft set theory has moved from a purely conceptual model of vagueness into a solid framework for association rule mining in complex, uncertain environments. Foundational works showed how transactional data can be transformed into soft information systems so that support, confidence, and maximal rules are defined directly over parameters, with maximal soft rules matching rough-set and classical results but at much lower runtime. Later studies refined this foundation with efficient maximal-rule mining structures (e.g., tree-based schemes, CRS pruning, clustering-assisted preprocessing) and extended soft sets into richer variants such as fuzzy soft sets, soft–fuzzy sets, temporal soft sets, N-soft sets, and pliancy/tree soft sets. These variants make it possible to handle real-valued, noisy, temporal, multi-group, and hierarchical data while keeping rule semantics interpretable. Empirical applications span student skill profiling and promotion strategy, maritime logistics risk early warning, heart-disease decision support, text and web usage mining, and amino-acid association analysis for TB and dengue, all of which demonstrate that soft-set-based ARM can capture compact, domain-aligned patterns that are difficult to obtain with purely crisp or black-box approaches.

Future research should focus on turning these conceptual and mid-scale successes into robust, large-scale, and fully hybrid systems. Key directions include: (i) scalable implementations of soft-set-based ARM for big and streaming data (e.g., distributed or online algorithms for temporal and N-soft sets); (ii) automatic learning of parameters, thresholds, and weights, rather than relying on manual tuning; (iii) concrete hybrid models that combine soft sets with fuzzy/rough sets, genetic algorithms, and representation learning (for example, GA-optimized fuzzy-soft association rules over temporal N-soft sets); (iv) richer temporal and spatio-temporal modeling to support long-term health monitoring, logistics, and IoT/sensor applications; and (v) systematic empirical benchmarks against modern ML and pattern-mining baselines, embedded in real decision-support systems in education, healthcare, logistics, and bioinformatics. Together, these directions can close the gap between elegant theory and production-ready, interpretable soft-set-driven rule engines.

References

[1] F. Fatimah, S. A. A. Kharis, and F. I. Fajar, “N-Soft Sets Association Rule and Its Application for Promotion Strategy in Distance Education,” Barekeng: Jurnal Ilmu Matematika dan Terapan, vol. 18, no. 3, pp. 1865–1878, Jul. 2024, doi: 10.30598/barekengvol18iss3pp1865-1878.

[2] D. Rohidin, N. A. Samsudin, and M. M. Deris, “Association Rules of Fuzzy Soft Set Based Classification for Text Classification Problem,” Journal of King Saud University – Computer and Information Sciences, vol. 34, no. 3, pp. 801–812, Mar. 2022, doi: 10.1016/j.jksuci.2020.03.014.

[3] X. Jia and D. Zhang, “Prediction of Maritime Logistics Service Risks Applying Soft Set Based Association Rule: An Early Warning Model,” Reliability Engineering & System Safety, vol. 207, p. 107339, Mar. 2021, doi: 10.1016/j.ress.2020.107339.

[4] G. S. Hussein, K. A. Eldrandaly, A. N. H. Zaied, S. L. Elhawy, and M. Mohamed, “Harnessing Pliancy Tree Soft Sets in Heart Diseases for Extracting Beneficial Association

Rules," International Journal of Intelligent Systems and Applications, vol. 13, no. 2, pp. 45–56, 2021.
[5] X. Liu, F. Feng, Q. Wang, R. R. Yager, H. Fujita, and J. C. R. Alcantud, "Mining Temporal Association Rules with Temporal Soft Sets," Journal of Mathematics, vol. 2021, pp. 1–17, Nov. 2021, doi: 10.1155/2021/7303720.
[6] U. Chandrasekhar, G. Vaishnavi, and D. Lakshmi, "Mining Maximal Association Rules on Soft Sets Using Critical Relative Support Based Pruning," in Lecture Notes on Data Engineering and Communications Technologies, Cham, Switzerland: Springer, 2020, pp. 931–940, doi: 10.1007/978-3-030-32150-5_94.
[7] F. Feng, Q. Wang, R. R. Yager, J. C. R. Alcantud, and L. Zhang, "Maximal Association Analysis Using Logical Formulas Over Soft Sets," Expert Systems with Applications, vol. 159, p. 113557, Nov. 2020, doi: 10.1016/j.eswa.2020.113557.
[8] P. D. Lanjewar and B. F. Momin, "A Review: Parameter Reduction Methods in Soft Set Theory," in Proc. 2019 Int. Conf. Communication and Electronics Systems (ICCES), 2019, pp. 715–722, doi: 10.1109/ICCES45898.2019.9002423.
[9] D. Rohidin, N. A. Samsudin, M. F. A. Aziz, and S. K. A. Khalid, "Using Fuzzy Soft Set Association Rule Mining Approach to Identify the Student Skill Data Association," Journal of Theoretical and Applied Information Technology, vol. 97, no. 13, pp. 3691–3701, 2019.
[10] A. Jain and K. R. Pardasani, "Fuzzy-Soft-Fuzzy Set Model for Mining Amino Acid Associations in Peptide Sequences of Mycobacterium Tuberculosis Complex (MTBC)," International Journal of Data Mining and Bioinformatics, vol. 17, no. 1, pp. 1–17, 2017, doi: 10.1504/IJDMB.2017.084021.
[11] A. Jain and K. R. Pardasani, "Fuzzy Soft Set Model for Mining Amino Acid Associations in Peptide Sequences of Mycobacterium Tuberculosis Complex (MTBC)," Journal of Intelligent & Fuzzy Systems, vol. 31, no. 1, pp. 259–273, Jun. 2016, doi: 10.3233/IFS-162139.
[12] F. Feng, M. Akram, B. Davvaz, and V. Leoreanu-Fotea, "Attribute Analysis of Information Systems Based on Elementary Soft Implications," Knowledge-Based Systems, vol. 70, pp. 281–292, Nov. 2014, doi: 10.1016/j.knosys.2014.07.010.
[13] M. I. Awang, A. N. M. Rose, F. Ahmad, M. M. Deris, and T. Herawan, "Discovering Attributes Dependency for Categorical Data Set Based on Soft Set Theory for Better Decision Making," Applied Mathematics & Information Sciences, vol. 9, pp. 6477–6490, 2015, doi: 10.12988/ams.2015.54360.
[14] B. Li, Z. Pei, and K. Qin, "Association Rules Mining Based on Clustering Analysis and Soft Sets," in Proc. IEEE Int. Conf. Computer and Information Technology, 2015, pp. 675–680, doi: 10.1109/cit/iucc/dasc/picom.2015.97.
[15] A. Jain and K. R. Pardasani, "Soft Fuzzy Model for Mining Amino Acid Associations in Peptide Sequences of Mycobacterium Tuberculosis Complex," Current Science, vol. 110, no. 4, pp. 603–618, Feb. 2016, doi: 10.18520/cs/v110/i4/603-618.
[16] S. A. H. Minoofam, J. Ahmadi, and H. R. Kanan, "A Comparative Review on Nondeterministic Sets for Association Rule Mining," in Proc. 2015 Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), 2015, pp. 1–5, doi: 10.1109/CFIS.2015.7391691.

[17] B. Vo, N. L. Minh, T. Tran, and T.-P. Hong, "Using Soft Set Theory for Mining Maximal Association Rules in Text Data," Journal of Universal Computer Science, vol. 22, no. 6, pp. 802–822, 2016, doi: 10.3217/JUCS-022-06-0802.

[18] T. Herawan, I. T. R. Yanto, and M. Mat Deris, "Soft Set Approach for Maximal Association Rules Mining," Database Theory and Application, vol. 64, pp. 163–170, 2009, doi: 10.1007/978-3-642-10583-8_19.

[19] T. Herawan and M. M. Deris, "A Soft Set Approach for Association Rules Mining," Knowledge-Based Systems, vol. 24, no. 1, pp. 186–195, Feb. 2011, doi: 10.1016/j.knosys.2010.08.005.

[20] D. Molodtsov, "Soft Set Theory—First Results," Computers & Mathematics with Applications, vol. 37, no. 4–5, pp. 19–31, Feb. 1999, doi: 10.1016/S0898-1221(99)00056-5.

[21] R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules Between Sets of Items in Large Databases," in Proc. ACM SIGMOD Int. Conf. Management of Data, 1993, pp. 207–216, doi: 10.1145/170035.170072.

[22] P. Bhatia, Data Mining and Data Warehousing: Principles and Practical Techniques. Cambridge, U.K.: Cambridge Univ. Press, 2019.

[23] D. Rohidin, N. A. Samsudin, and T. Herawan, "On Mining Association Rules of Real-Valued Items Using Fuzzy Soft Set," in Advances in Intelligent Systems and Computing, Cham, Switzerland: Springer, 2017, pp. 517–528, doi: 10.1007/978-3-319-51281-5_52.

[24] I. T. R. Yanto and A. Rahman, "Soft Maximal Association Rule for Web User Mining," International Journal of Computer Applications, vol. 140, no. 10, pp. 1–6, 2016.

[25] A. Gour and K. R. Pardasani, "Soft Fuzzy Set Approach for Mining Frequent Amino Acid Associations in Peptide Sequences of Dengue Virus," Proceedings of the National Academy of Sciences, India Section A: Physical Sciences, vol. 88, no. 4, pp. 529–538, Dec. 2018, doi: 10.1007/s40010-016-0336-3.

[26] S. R. Dash and S. Dehuri, "Soft Set and Genetic Algorithms for Association Rule Mining: A Road Map and Direction for Hybridization," in Advances in Intelligent Systems and Computing, vol. 199, Berlin, Germany: Springer, 2013, pp. 381–393, doi: 10.1007/978-3-642-35314-7_44.

[27] S. R. Dash, S. Dehuri, and U. K. Sahoo, "Interactions and Applications of Fuzzy, Rough, and Soft Set in Data Mining," International Journal of Fuzzy Systems and Applications, vol. 3, no. 3, pp. 37–50, Jul. 2013, doi: 10.4018/ijfsa.2013070102.