A Study of Algorithms for Mining Fuzzy Association Rules

Nguyễn Bình Minh
Thai Nguyen University of Information and Communication Technology
Thai Nguyen University, Thai Nguyen, Vietnam

***Abstract*—This paper presents an English IMRAD summary of a 2007 undergraduate thesis on fuzzy association-rule mining. The thesis extends conventional association rules to quantitative attributes represented by fuzzy sets, describes an Apriori-style candidate-generation and pruning procedure, and implements an experimental tool named FuzzyARM in Microsoft Visual C#. It examines minimum fuzzy support and confidence, the number of fuzzy attributes and records, and the choice of triangular norm. In the reported experiments, reducing minimum support from 15% to 1% increased processing time from 0.09 to 24.08 seconds; increasing fuzzy attributes from 13 to 22 increased time from 0.031 to 26.844 seconds. For 271 to 138,241 records, the reported times ranged from 0.063 to 2.375 seconds. The thesis concludes that algebraic product was the most suitable of the tested triangular norms for its data and rule-quality judgment. Hardware details, repeated runs, and an external benchmark are not reported, so these numbers should be read as descriptive results of the original experiment.**

***Index Terms*—fuzzy association rules, data mining, Apriori, fuzzy support, triangular norm, FuzzyARM**

I. Introduction

Association-rule mining identifies relationships among items in transaction data [2], [3]. Conventional rules are less direct when a record contains numerical values such as age or blood measurements. Discretization can discard graded membership near a boundary. The source thesis studies fuzzy association rules as a way to represent such attributes using membership functions and then mine rules with support and confidence thresholds [1], [4].

The objective is to explain the mathematical and algorithmic basis of the approach, implement a small experimental tool, and examine how thresholds and data dimensions affect rule generation and execution time. The thesis draws on general data-mining texts [5], [6], [7] and presents its own Apriori-like procedure and examples.

II. Methods

A. Fuzzy rule representation

The thesis replaces a crisp numerical interval with a fuzzy set and its membership function. A triangular norm combines membership values when an itemset contains multiple fuzzy attributes. The proposed workflow first builds frequent one-attribute sets, joins them to form candidate sets, prunes candidates whose subsets are not frequent, checks fuzzy support against a minimum threshold, and generates rules whose confidence exceeds a second threshold [1]. Candidates containing two fuzzy sets derived from the same original attribute are excluded.

Four triangular norms are discussed: minimum, algebraic product, bounded product, and drastic product. The thesis also proposes thresholding fuzzy membership values to produce a comparable numerical-attribute rule. Its worked example converts a fuzzy age-and-blood-sugar rule to a numerical statement while retaining the reported support and confidence [1].

B. Implementation and evaluation

The author implemented FuzzyARM in Microsoft Visual C# and reported experiments using cardiovascular-diagnosis, diabetes-diagnosis, and transport datasets [1]. The evaluation varied minimum fuzzy support, minimum confidence, the number of attributes, the number of records, and the triangular norm. Other settings were held fixed within each reported comparison. The thesis presents times, counts of frequent sets, and counts of accepted rules, but does not provide machine specifications or statistical uncertainty.

III. Results

The number of reported frequent sets rose from 26 at 30% minimum support to 833 at 1%; the number of accepted rules rose from 2 to 665 over the same support range. In a separate confidence-threshold experiment, accepted rules rose from 24 at 100% minimum confidence to 968 at 60% [1]. These are different parameter sweeps and should not be combined into a single comparison.

The source reports substantial variation among triangular norms. Minimum and drastic product generated more accepted rules than algebraic and bounded product in its evaluation. It favored algebraic product because the bounded-product rules were judged less useful, while a larger number of rules alone did not indicate better rule quality [1].

TABLE I
Selected measurements reported in the source thesis [1].

Changed setting	Low end	High end
Min. support, 15% to 1%	0.09 s	24.08 s
Fuzzy attributes, 13 to 22	0.031 s	26.844 s
Records, 271 to 138,241	0.063 s	2.375 s

Supervisor: Vu Vinh Quang, Thai Nguyen University of Information and Communication Technology (vvquang@ictu.edu.vn).

IV. Discussion

The observed growth in processing time as minimum support falls is consistent with the greater number of candidate and frequent itemsets. The increase with additional fuzzy attributes is particularly marked in the reported measurements. The record-count experiment spans a larger range but shows a smaller increase in time; the thesis describes this pattern as approximately linear [1]. These observations are descriptive because the document does not give hardware, repeated-trial variation, or a matched comparison with another implementation.

Fuzzy representation can express graded boundaries in quantitative data, but the selected membership functions, thresholds, and triangular norm shape the resulting rules. The thesis's practical recommendation of algebraic product therefore applies to its examples and qualitative assessment, not to every dataset. A follow-up evaluation would need explicit rule-quality criteria and reproducible datasets before making a broader claim.

V. Conclusion

The undergraduate work explains and implements an Apriori-style method for fuzzy association rules and demonstrates the FuzzyARM tool. Its experiments document sensitivity to support, confidence, data dimensions, and triangular norm. The reported values provide useful historical evidence of the implementation, while the absence of a controlled external baseline limits comparative conclusions [1].

References

[1] N. B. Minh, "A study of algorithms for mining fuzzy association rules," undergraduate thesis (original in Vietnamese), Thai Nguyen University of Information and Communication Technology, 2007. [Online]. Available: ICTU repository record.

[2] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," *ACM SIGMOD Record*, vol. 22, no. 2, pp. 207–216, 1993, doi: 10.1145/170036.170072.

[3] H. Mannila, H. Toivonen, and A. I. Verkamo, "Efficient algorithms for discovering association rules," in *KDD-94: AAAI Workshop on Knowledge Discovery in Databases*, 1994, pp. 181–192.

[4] A. Gyenesei, "A fuzzy approach for mining quantitative association rules," TUCS Technical Report 336, Turku Centre for Computer Science, 2000.

[5] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. San Francisco, CA, USA: Morgan Kaufmann, 2005.

[6] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. San Francisco, CA, USA: Morgan Kaufmann, 2002.

[7] A. Rea, "Data mining: An introduction," The Parallel Computer Centre, Queen's University of Belfast, 1995.

[8] Đ. M. Tường, *Artificial Intelligence* (in Vietnamese). Hanoi, Vietnam: Vietnam National University Press, 2003.

[9] H. V. Kiếm, *How to Solve a Problem on a Computer*, vol. 3 (in Vietnamese). Hanoi, Vietnam: Education Publishing House, 2004.

[10] P. Đ. Diệu, *Logic in Knowledge Systems* (in Vietnamese). Hanoi, Vietnam: Vietnam National University Press, 1999.