International Journal of Advance and Applied Research
www.ijaar.co.in
ISSN – 2347-7075 Impact Factor – 8.141
Peer Reviewed Bi-Monthly
Vol. 6 No. 22 March - April - 2025

Redefining Interestingness Measures Using Fuzzy Support Matrix in Data Mining

Dr. Swati Joshi
Pune Vidyarthi Griha's College of Science and Commerce, Pune
Corresponding Author – Dr. Swati Joshi
DOI - 10.5281/zenodo.15533013

Abstract:

Interestingness measures plays a crucial role in evaluating the relevance and significance of patterns discovered in data mining. Traditional way of defining interestingness measures rely on rigid, deterministic thresholds. In real world data inherent uncertainty and vagueness is present and the interestingness measures often fail to capture the. In this paper we introduce a novel approach for redefining interestingness measures using a Fuzzy Support Matrix (FSM). Integration of fuzzy logic to evaluate the interestingness, for offering more flexible and nuanced measure of pattern relevance. It helps in robust decision-making especially in uncertain and complex environments. This research work propose new fuzzy-based interestingness metrics. This paper shows analysis of the advantages of newly proposed methods over traditional methods, and also demonstrate their effectiveness through case studies in various domains, such as e-commerce, healthcare, and social networks.

Introduction:

Data mining uncovers valuable patterns from large and complex datasets. However, not all patterns discovered during the mining process are interesting and useful to the end-user. Hence there is need for interestingness measures that can evaluate the quality of patterns in terms of their relevance and significance. Traditionally interestingness measures, such as support, confidence, and lift, are frequently used which are based on deterministic thresholds and hence may not reflect the uncertainty and subjectivity involved in real-world data mining. In this paper, we propose redefining interestingness measures by incorporating fuzzy logic into the evaluation process. Specifically, we introduce the concept of a Fuzzy Support Matrix (FSM) to represent uncertainty and vagueness in the support and confidence of patterns. By doing so, we aim to develop more flexible and meaningful interestingness measures that are better suited for complex, noisy, and imprecise datasets.

Motivation:

Traditionally defined interestingness measures are insufficient in handling uncertain or imprecise data, which is very common in many real-world applications. Fuzzy logic have ability to model uncertainty and vagueness, enhancing the evaluation of patterns, making it more adaptable to various domains and situations.

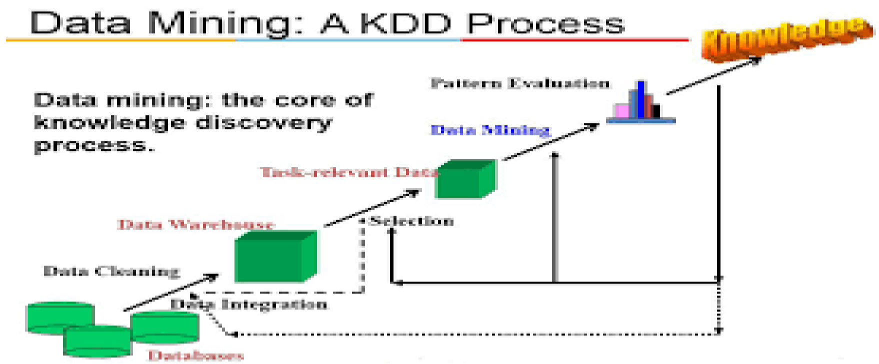
Objectives:

1. Introducing the concept of a Fuzzy Support Matrix in a different way as a tool for

redefining interestingness measures.
2. Propose new fuzzy-based interestingness metrics to handle uncertainty and vagueness.
3. Comparative analysis of fuzzy-based interestingness measures with traditional ones in terms of effectiveness and flexibility.
4. Showcase the applicability of fuzzy interestingness measures in real-world data.

Background and Related Work:

Data mining techniques are used to extract useful knowledge from large datasets. Some of the data mining methods are association rule mining, classification, clustering, and anomaly detection. Association rule mining, uses the interestingness measures such as support, confidence, and lift to evaluate the quality of discovered rules. However, these measures are limited because they do not account for the inherent uncertainty or fuzziness present in real-world data.



Traditional Interestingness Measures:

Traditional interestingness measures gives the statistical properties of patterns and filters irrelevant patterns. The most commonly used measures are:
Support: Measures the frequency of the pattern in the dataset.
Confidence: Measures the probability that an item appears in a transaction given the presence of another item.
Lift: Measures the ratio of the observed frequency of a pattern to the expected frequency.

Traditional measures are inadequate if the data contains noise, vagueness, or imprecision, which is common in domains such as healthcare or social media analytics.

Fuzzy Logic in Data Mining:

Fuzzy logic has been used to handle uncertainty and imprecision in data mining. In fuzzy systems, instead of representing data in binary terms (true or false), elements are assigned membership degrees between 0 and 1, that allows for more flexibility in modeling uncertainty. Many researchers have explored applying fuzzy logic to data mining tasks, such as clustering, classification, and association rule mining. However, there are very less attempts on applying fuzzy logic to interestingness measures.

Fuzzy Support Matrix:

Many researchers have used 2 X 2 contingency table to define interestingness mea- sures ([Tan et al. (2004)], [Zembowicz and Z˙ ytkow (1996)], [Lenca et al. (2007)], [Yao and Zhong (1999)], [Xuan-hiep et al. (2006)]). The table actually stores frequenc or support count of sets

$A\bar{B}$, $\bar{A}B$, $\bar{A}\bar{B}$ and AB. These support counts can be easily computed for binary attributes as it is either present or absent. However for quantified or categorical attributes fuzzy approach can be used to handle this vagueness. A small or negligible value of quantified attribute may define absence and a high or substantial value may represent strong presence. For attributes A and B the fuzzy association rule ALow → BLow indicates when"A is in low quantity so is B" and is equivalent to negative implication of $\bar{A} \rightarrow \bar{B}$. The support count of $\bar{A} \rightarrow \bar{B}$ can be replaced by fuzzy support count of ALow → BLow. The fuzzy association rule ALow → BHigh implies when A is in low quantity then B is in high quantity, thus approximates $\bar{A} \rightarrow B$ hence support count of $\bar{A} \rightarrow B$ can be replaced by fuzzy support count of ALow → BHigh. Similarly the fuzzy support of AHigh → BLow approximates $A \rightarrow \bar{B}$. The fuzzy association rule AHigh → BHigh implies A is high so is B indicating positive relationship A → B. The support matrix is defined for two attributes A and B. Each attribute is partitioned into m fuzzy partitions {F 1, F 2, F 3, ...Fm} m≥2 where F 1 indicates Low and Fm indicates high giving rise to fuzzy sets {F 1 , F 2 , F 3 , ...Fm}, which for attribute A are equivalent to {FLow, ..., FHigh}. The fuzzy sets for attribute B will be {FLow, ..., THigh}.

Let S^{AB} denote fuzzy support count for $A_{Low} \rightarrow B_{High}$, where L and H are abbreviations for Low and High respectively. The term SAB indicates support count when attribute A is Low while attributes B can take any possible value from low to high. The term S^{AB} indicates support count when attribute A takes any possible values while attribute B is low. Similar interpretation can be given for the terms S^{AB} and S^{AB}. The support matrix is shown in table 1.

Table 1. Fuzzy Support Matrix

Attributes	B_{Low}		B_{High}	
A_{Low}	S^{AB}_{LL}	....	S^{AB}_{LL}	S^{AB}_{L+}
....	....	....	....	.....
A_{High}	S^{AB}_{HL}	....	S^{AB}_{HH}	S^{AB}_{H+}
	S^{AB}_{+L}	....	S^{AB}_{+H}	n_f

Redefining Interestingness Measures Using Fuzzy Support Matrix:

We propose the following fuzzy-based interestingness metrics: In defining different interestingness measures, four cell values from the support matrix were used namely s^{AB}_{LL}, s^{AB}_{LH}, S^{AB}_{HL}, and S^{AB}_{HH}. The first and last values represent the extent to which two attributes complement each other, while middle values represent how they contradict or oppose each other. Following table shows the different redefined interestingness measures.

Sr.No	Interestingness Measures	Definition
		Support Matrix
1	Support	$\frac{S_{HH}^{AB}}{n_F}$
2	Confidence/Precision	$\frac{S_{HH}^{AB}}{S_{H+}^{AB}}$
3	Coverage	$\frac{S_{H+}^{AB}}{n_F}$
4	Prevalence	$\frac{S_{+H}^{AB}}{n_F}$
5	Recall	$\frac{(S_{HH}^{AB})}{(S_{+H}^{AB})}$
6	Specificity	$\frac{(S_{LL}^{AB})}{(S_{L+}^{AB})}$
7	Accuracy	$\frac{(S_{HH}^{AB})}{n_F} + \frac{(S_{LL}^{AB})}{n_F}$
8	Lift/Interest	$n_F * \frac{S_{HH}^{AB}}{S_{H+}^{AB} S_{+H}^{AB}}$
9	Leverage	$\frac{S_{HH}^{AB}}{S_{H+}^{AB}} - \frac{S_{H+}^{AB}}{n_F} * \frac{S_{+H}^{AB}}{n_F}$
10	Added Value	$\frac{S_{HH}^{AB}}{S_{H+}^{AB}} - \frac{S_{+H}^{AB}}{n_F}$

Advantages of Fuzzy Interestingness Measures:

- **Flexibility:** Fuzzy interestingness measures provide more flexibility than traditional deterministic measures, accommodating varying degrees of support and confidence.
- **Handling Uncertainty Effectively:** Incorporating fuzzy logic, interestingness measures can better handle uncertainty, vagueness, and imprecision in the data.
- **Domain Adaptability:** Fuzzy interestingness measures can be adapted to specific domains where data is inherently uncertain or where precise thresholds are difficult to define.

Evaluation and Comparison:

	Measure Values for Rules					
Measures /Rules	SL→SW	SL→PL	SW→PL	SL→PW	SW→PW	PL→PW
Support	0.06	0.36	0.33f	0.0801	0.0841	0.5007
Confidence	0.1698	1	1	0.2194	0.2318	1
Coverage	0.3662	0.366	0.3393	0.3651	0.3630	0.5007
Prevalence	0.2912	0.3667	0.3393	0.3896	0.3876	0.5007
Recall	0.2135	1	1	0.2056	0.2171	1
Specificity	0.4088	1	1	0.0785	0.0691	1
Accuracy	0.1812	0.9371	0.9016	0.1047	0.1056	1.1428
Lift	0.5829	2.7266	2.9466	0.5632	0.5980	1.9969
Leverage	0.0631	0.8654	0.8848	0.0771	0.0910	0.7492
Added Value	-0.121	0.6332	0.6606	-0.1701	-0.1558	0.4992

We evaluated an experiment on different data sets for the performance comparison. Results are shown in above table for the experiment which we carries out on IRIS data set. The same can be applied on various domains like E- commerce, healthcare and social network further. In E-commerce it can be used for analyzing patterns in customer purchasing behavior and product recommendations. In healthcare to identify patterns in patient data to predict disease outbreaks or treatment effectiveness. In Social Networks for discovering patterns in user behavior, social connections, and emerging trends.

Results:

The fuzzy support matrix has been used for defining various interestingness measures in a uniform and consistent manner. The extended definition of existing measures are applied to identify new attribute characteristics. Theoretical evaluation of various properties for interestingness measures has been provided both in structural and behavioral aspects. Experimental study of these measures on different datasets is done to strengthen their importance. The fuzzy-based measures outperform over traditional ones in terms of adaptability and robustness, especially in domains where data is noisy or vague.

Conclusion:

This paper introduces a novel approach to redefining interestingness measures by incorporating fuzzy logic through the use of a Fuzzy Support Matrix (FSM). By leveraging fuzzy membership functions, we have proposed more flexible, adaptable, and meaningful interestingness metrics that can handle uncertainty and vagueness in data. The proposed fuzzy-based metrics are evaluated through case studies and shown to be effective in various domains. Future work should focus on optimizing fuzzy logic systems and exploring their integration with other advanced data mining techniques.

References:

1. Zadeh, L. A. (1965). Fuzzy sets. Information and Control, 8(3), 338–353.
2. Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), 487–499.
3. Klir, G. J., & Folger, T. A. (1999). Fuzzy Sets, Uncertainty, and Information. Prentice Hall.
4. Swati R. Ramdasi; Shailaja C. Shirwaikar AICTC '16: Proceedings of the International Conference on Advances in Information Communication Technology & Computing, Article No.: 88, Pages 1 – 5
5. Bayardo Jr, R. J. and R. Agrawal (1999). Mining the most interesting rules. In Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 145–154. ACM.
6. Brin, S., R. Motwani, and C. Silverstein (1997). Beyond market baskets: Generalizing association rules to correlations. In Acm Sigmod Record, Volume 26, pp. 265–276. ACM.
7. Swati R. Ramdasi; Shailaja C. Shirwaikar; Vilas Kharat , Interestingness measures for quantified and ordered categorical attributes using fuzzy approach International Journal of Fuzzy Computation and Modelling (IJFCM), Vol. 2, No. 4, 2019
8. Chan, K. C. and W.-H. Au (1997). Mining fuzzy association rules. In Proceedings of the sixth international conference on Information and knowledge management, pp. 209–215. ACM.
9. Dieu, P. D. Logic in knowledge systems. Faculty of Technology, Hanoi National. Dougherty, J., R. Kohavi, M. Sahami, et al. (1995). Supervised and unsupervised discretization of continuous features. In Machine learning: proceedings of

the twelfth international conference, Volume 12, pp. 194–202.

10. Dubois, D., E. H¨ullermeier, and H. Prade (2003). A note on quality measures for fuzzy association rules. In International Fuzzy Systems Association World Congress, pp. 346–353. Springer.
11. H., G. Fabrice, and B. Henri (2006). Discovering the stable clusters between interestingness measures. Yang, Y. and M. Singhal (1999). Fuzzy functional dependencies and fuzzy association rules. In International Conference on Data Warehousing and Knowledge Discovery, pp. 229–240. Springer.
12. Yao, Y. and N. Zhong (1999). An analysis of quantitative measures associated with rules. In Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 479–488. Springer.
13. Zadeh, L. A. (1965). Fuzzy sets. Information and control 8 (3), 338–353. Zembowicz, R. and J. M. Zytkow (1996). From contingency tables to various forms of knowledge in databases. In Advances in knowledge discovery and data mining, pp. 328–349.