DHĀRANA

INTERNATIONAL JOURNAL OF BUSINESS FROM BHARATIYA VIDYA BHAVAN'S
M. P. BIRLA INSTITUTE OF MANAGEMENT, BENGALURU
Vol.11, #2 (2017) pp 13-20 ISSN 0974-0082

Issues in Negative Association Rule Mining with Business Analytics Perspectives

Rupal Sethi[1] and B. Shekar[2]

Abstract

Association Rule mining literature is witnessing a shift of focus from generating positive rules to the discovery of negative rules. A review of previous literature on negative rule mining that incorporate objective and subjective interestingness measures has been done. Then, an extension, to Fuzzy Set Concept for generating and mining negative rules is made. This work also presents unaddressed issues in mining of both positive and negative rules. Business applications that gain useful insights from both positive and negative rules have been highlighted.

Keywords: *Association Rule Mining; Item sets, Negative Association Rules; Fuzzy Set Concept; Interestingness; Business Applications.*

Introduction

An Association Rule (AR) is an implication of the form "$X \rightarrow Y$", where I is a set of items and $X \subseteq I$; $Y \subseteq I$, and $X \cap Y = \phi$. One such rule would be: Bread → Butter. This rule says customers buying bread are likely to buy butter as well. AR Mining has been applied to broadly two types of data: transaction data and quantitative attribute data. The transaction datasets comprise items that are associated together through an event such as market basket or web log analysis. For example, the rule Bread, *Jam → Butter* is obtained from transaction data. Quantitative attribute data consists of variables that are either binary or categorical in nature. Quantitative association rules are generated by partitioning these categorical or binary variables (Srikant and Agrawal 1996). One such example would be *<Age=30-40>, <Gender=Female> → <No. of cars=2>.*

Contrary to positive AR, there are negative rules as another category of ARs (Brin, Motwani & Silverstein 1997). These depict relationship between items that are in conflict like people who buy Pepsi do not buy Coke. Thus, negative ARs identify items that a customer

[1] *Research Scholar, Fellow Programme in Management, IIMB, Bangalore 560 076.*
[2] *Professor, Decision Sciences and Information Systems Area, Indian Institute of Management Bangalore, Bangalore, 560 076, India. emailLshek@iimb.ernet.in*

is not likely to buy whenever he buys a certain set of items (Savasere, Omiesinski & Navathe 1998). A typical negative AR is represented as "X → ~Y". Similar to positive AR, negative AR can also be formed using both transaction data *(Baby Soap → ~ Facewash)* and quantitative attribute data *(<Age=20-30>, <Married=Yes> → ~ <Days of purchase=Weekday>).*

Generation of quantitative ARs requires partitioning of attributes. Partitioning of categorical variables often leads to information loss. In order to minimize this information loss, fuzzy set concept is used in the literature (Kuok, Fu and Wong 1998). However, past work is limited to applying fuzzy set concept in pruning positive ARs. This paper addresses the issues pertaining to generating negative rules using fuzzy sets. An attempt to erudite various business applications that may benefit from such analysis is elucidated.

Organisation of the paper: In Section 2 we present the previous literature on negative association rules. State-of-the-art objective and subjective interestingness measures that are used for generating negative rules are described. The Fuzzy set concept for positive ARs in Section 3 and expand on it for mining negative ARs is provided. Throughout, some issues that crop up in mining negative rules are highlighted. In Section 4, three major business applications that benefit from using knowledge gained from mining positive and negative ARs: Market Basket Analysis, Customer Relationship Management and Credit Scoring are given.

2 Negative Association Rules

2.1 Background and Motivation

The concept of negative association rules is still nascent in the field of data mining. There have been some attempts to develop algorithms for generating negative association rules. The discovery process is a difficult task since the search space for negative rules is too large. Absence of itemsets cannot be programmed and even if it is, that leads to generation of millions of negative rules that may not be of use to the manager. Hence, the objective is to find only "interesting" negative association rules that can be acted upon by the managers. The de facto interestingness measures used for generating positive association rules are *support* and *confidence* (Agrawal and Srikant 1994). These measures in the Apriori algorithm (Agrawal and Srikant 1994) prune the itemsets based on the threshold for frequency count. Unlike positive association rules, negative rules cannot be generated by a Apriori algorithm since they involve absence of items. Thus, researchers have modified the Apriori algorithm for negative rules using different interestingness measures like correlation and expectation. There have been attempts to use subjective interestingness measures like unexpectedness for generating negative rules.

The inception of the idea of negative implications was given by Brin, Motwani and Silverstein (1997). They extended the support-confidence framework by necessitating the use of correlation coefficient in generating interesting rules. They argued that support and confidence cannot highlight the negative relationship between two sets of items, while correlation gives the strength as well as direction of relationship. Consider the example of milk and jam.

The following transaction frequency table has been made from a hypothetical transaction dataset:

	Jam	**No Jam**	
Milk	5	4	9
No Milk	6	3	9
(Total)	**11**	**7**	**36**

Table 1: Frequency table for milk and jam

Consider the itemset: (Milk, Jam)

We generate rules using Apriori with thresholds minsupp= 10%, minconf= 50%

$$Support = P(Milk\ and\ Jam) = \frac{5}{36} = 0.14$$

$$Confidence = \frac{P(Milk\ and\ Jam)}{P(Milk)} = \frac{5}{9} = 0.56$$

Since *support* and *confidence* of this itemset are greater than the thresholds, the rule *Milk* → *Jam* is generated.

However, if we calculate the correlation coefficient between milk and jam, we get an altogether different

$$Correlation = \frac{P(Milk\ and\ Jam)}{P(Milk)P(Jam)} = \frac{5}{(9)(11)} = 0.05$$

Thus, milk and jam are indeed negatively correlated implying that people, who buy milk, *do not* buy jam. This relationship is not captured in the support-confidence framework. Thus the rule *Milk* $\rightarrow$ *Jam* is misleading in the absence of information on correlation between itemsets.

2.2 Objective Interestingness Measures

2.2.1 Correlation and Lift

The concept of negative correlation by Brin et al (1997a) led to a stream of research in negative association rules. However, the authors do not use the original measure of correlation (Pearson 1895) but rather use *lift* (Brin, Motwani, Ullman and Tsur 1997b) as a proxy to it. One of the problems with lift is that it doesn't consider the complement forms of itemsets. As a result of which, negative rules of the form $A \rightarrow \sim B$ *or* $\sim A \rightarrow B$ cannot be formed.

The work by Brin et al (1997a) on negative implications was extended by Antonie and Zaiane (2004). They used Pearson's correlation coefficient as a measure of negative association. They provide an algorithm that extends the support-confidence framework from the Apriori with a sliding correlation coefficient threshold. The algorithm checks for minimum support and confidence first, and then checks for correlation. If the correlation is positive and greater than a threshold (t), positive ARs are generated. On the other hand, if the correlation is negative and greater than the threshold in magnitude, negative ARs of the form $A \rightarrow \sim B$ or $\sim A \rightarrow B$ are generated.

The algorithm used by Antonie and Zaiane (2004) generates both positive and negative association rules using a single threshold value for support, confidence and correlation. Although, this approach saves time and space, we are not sure if it still generates *interesting* negative rules. Also, even if the algorithm is able to generate both types of rules, one fails to see the link between positive and negative rules. Both the types are independent of each other and no common knowledge seems to emerge from them.

Another issue with this algorithm is that it is restricted to just two items, one as an antecedent and other as a consequent. This assumption makes the problem quite naïve. If we consider more than two items in the itemset for candidate sets, we need to establish the correlation between different combinations of items. For example, consider (milk, jam, butter) as a candidate itemset. For a rule to be generated from this itemset, one need to have a positive correlation between items falling on either side.

If the negative rule is *Milk, Butter* $\rightarrow \sim$ *Jam*

Milk and Butter should have a high positive correlation and (Milk, Butter) should have a high negative correlation with Jam. Considering this case of correlation among and across itemsets, one single value for correlation threshold might not suffice.

So far, negative rules have been generated using two objective measures: lift and Pearson's correlation coefficient. As stated earlier, Brin et al (1997a) used lift as a proxy for correlation. The reason for this substitution has not been cited by the authors but considering the complexity of Pearson's coefficient, one can understand the conceptual notion the paper tried to bring in. However, we should realize the difference between both the measures in assessing negative relationships. Also, one can look at other better measures that can be used for identifying negative relationships between items.

$$Lift = \frac{P(AB)}{P(A)P(B)}$$

$$Pearson's Correlation\ \rho = \frac{Covariance(A,B)}{sd(A)\ \ sd(B)}$$

On simplification,

$$\rho = \frac{P(AB) - P(A)P(B)}{\sqrt{P(A)\ P(B)P(\overline{A})\ P(\overline{B})}}$$

Focusing on numerators, Lift only takes into account the frequency of occurrence of both items together, while Pearson's correlation calculates the difference between co-occurrence and independent occurrence. Thus, Pearson's correlation coefficient gives a proper measure of negative relationships.

2.2.2 Alternate Interestingness Measure

Since we are aware that Pearson's correlation coefficient is complex in calculations because of the denominator, we provide alternative measure of objective interestingness that captures the negative relationship similar to Pearson's coefficient. *Change of Support* (CS) is a measure of interestingness that was formulated by Yao and Zhong (1999) but has not been studied that much by researchers in association rule mining.

$$CS = P(B/A) - P(B)$$

$$= \frac{P(AB) - P(A)P(B)}{P(A)}$$

Clearly, CS is similar to ρ as far as the numerator is concerned. CS is better in terms of complexity and intuition.

2.3 Subjective Interestingness Measure

Another approach to generating negative rules is by using the taxonomy of the dataset. A rule is interesting if it deviates from the manager's expectation based on previous belief. The previous belief is usually stated in terms of a priori probabilities based on knowledge of the problem domain (Savasere Omiecinski and Navathe 1998). This concept is termed as *unexpectedness.* The major assumption based on the taxonomy of the data is called the uniformity assumption. It states that items that belong to the same parent in a taxonomy are expected to have similar types of associations with other items. In other words, siblings in a taxonomy are *substitutable.* For example, if Chips are bought with Pepsi, one would expect Chips to be bought with Coke as well. If the actual support of Chips and Coke is less than the support of Chips and Pepsi, then Chips and Coke generate a negative association rule.

Savasere et al (1998) define negative rules as consisting of items that are not likely to be bought along with the purchase of a set of items. One of the conceptual questions we would like to pose here is the difference between the notion of not buying item A given the purchase of item B vis-à-vis the notion of buying item B decreasing the likelihood of buying item A. Although both notions look the same, there is a deeper meaning attached to the latter. The first definition can be applied to any set of two unrelated items A and B. However the second definition restricts to two related items. Consider the following:

Form 1: *Pen* → ~*Milk*

Form 2: *Tea* → ~ *Coffee*

According to form 1, a person is not likely to buy milk when he buys pen. Here, the rule makes perfect sense because pen and milk are quite unrelated. On the contrary, form 2 says that when a person is buying tea, he is less likely to buy coffee. The second form has a notion of causality as well as substitution. This paper talks about form 1 accompanied with the knowledge of taxonomy. Hence there are less chances of rule generation involving unrelated products. However one must be careful before giving a generic definition for negative rules.

The uniformity assumption made by Savasere et al (1998) stating that taxonomy consists of siblings that are substitutable, needs further probing. Two fundamental questions arise: 1) What do we mean by substitution here? and 2) Based on the application what is the level of granularity of the taxonomy that should be exercised?

Savasere et al (1998) do not define the meaning of substitution in the context of their paper. The concept of substitution should be linked to a function that specifies the level of information reusability. Substitution can be, at the level of brand (Colgate vs. Pepsodent Toothpaste), application-specific (flowers vs. chocolates), seasonality-driven (ice-cream vs. hot chocolate) or at a more abstract level (desktop vs. laptop). Thus, restricting to siblings can lead to over or under representation of negative rules. Savasere et al. do not restrict the uniformity assumption to siblings alone, but span across siblings, parents and children in three ways.

Buckles, Yuan and Zhang (2002) adopt an approach similar to that of Savasere et al (1998). They also use the concept of locality of similarity in defining sibling rules from the taxonomy. Sibling rules are a pair of positive association rules where both the siblings are expected to be related to the same consequent. For example, if *Pepsi* → *Chips* is a rule that is generated through Apriori, then *Coke* → *Chips* should also be generated. If the confidence of *Coke* → *Chips* is less than the expected confidence (equal to *Pepsi* → *Chips*), then a negative rule *Coke* → ~ *Chips* gets generated.

Domain knowledge being present in the taxonomy makes this also a subjective approach for generation of negative rules. Buckles et al (2002) assert that the criterion for a negative rule utility is its relationship to a valid positive rule. This idea of linking positive and negative rules through a taxonomy is indeed useful for managers for making strategic decisions. However, the same question arises here also - what level of substitution are we seeking? As one moves to greater abstract level up the hierarchy every item is substitutable by another. Hence, there needs to be a function defined for substitution of items that appear in negative rules. This paper is an attempt to extend the definition of substitution and use it for linking positive and negative association rules.

3 Fuzzy Set Approach

Kuok, Fu and Wong (1998) argue that fuzzy set concept is better than the discrete interval method (Srikant and Agrawal 1996) since it provides a smooth transition between member and non-member of a set while partitioning. Such an approach of fuzzy sets results in fewer boundary elements getting excluded while partitioning the quantitative attributes.

A positive AR is defined using fuzzy sets (Kuok et al 1998) as follows.

$$X \; is \; A \rightarrow Y \; is \; B$$

X and Y are quantitative attributes, and A and B are fuzzy sets corresponding to X and Y respectively.

Two interestingness measures that use fuzzy sets are employed for generating positive ARs. *Significance Factor* gives the number of records supporting the itemset and also their degree of support.

$$Significance < X, A > = \frac{Sum \; of \; the \; votes \; satis}{Total \; Number \; o}$$

Votes satisfying set <X,A> signifies the degree of membership of each record having attribute X lying in fuzzy class A. This measure is similar to *support* as it reflects the support for the itemset relative to the entire dataset

The interestingness measure is called *Certainty Factor.* Kuok et al (1998) use two methods to calculate *Certainty Factor,* but do not link the information gained from both the methods. We try to address this gap by separating the two methods as two distinct objective interesting measures for the positive AR.

The first method to calculate *Certainty Factor* uses *Significance.*

$$Certainty \; (X, A \rightarrow Y, B) = \frac{Significance \; of <}{Significance \; of <}$$

Here Z = XUY and C = AUB

This is similar to confidence for positive AR.

$$fidence \; (X \rightarrow Y) = \frac{Support \; of \; X \; U \; Y}{Support \; of \; X}$$

The second method uses Pearson's correlation coefficient. Since fuzzy rules are different from positive rules, calculation of expectation of antecedent and consequent is a little different. The vote of a record is zero if its membership function is less than a user specified threshold.

3.1 Generating Negative ARs using Fuzzy Sets

The two methods in Kuok et al (1998) present contrasting results for the same fuzzy rule. *Consider Salary, High* → *Balance, Low.* The *Certainty Factor* using *Significance* is positive (0.364) highlighting the fact that the consequent is 36.4% significant relative to the antecedent in the entire dataset. Contrary to this, *certainty factor,* (ϱ=-0.96), presents an entirely different picture. It shows that High Salary is strongly negatively correlated with low balance, implying that the rule should not be formed. Instead a negative rule should get generated *Salary, High* → ~ *Balance, Low.*

$$Confidence \; (X \rightarrow Y) = \frac{P(XY)}{P(X)}$$

$$Correlation \; (X \rightarrow Y) = \frac{P(XY) - P(X)P(Y)}{f(P(X)) \;\; f(P(Y))}$$

Hence, we argue that the second method is an additional measure of interestingness similar to the addition of correlation to the support-confidence framework by Brin et al (1997). This fuzzy correlation measure may be used to generate negative fuzzy rules. We also recommend that the thresholds used for both *Certainty factors* should be different as they give different relationships between itemsets. Identical thresholds, as mentioned by Kuok et al. may generate misleading rules.

We summarise our observations on *Certainty factor.*

i. *Certainty factor* based on *significance* and that based on *Correlation* portray different relationships between itemsets. Hence they should be considered as two different objective interestingness measures for fuzzy rules.

ii. *Significance*-based *Certainty factor* is similar to *confidence* as it measures the support of consequent relative to the antecedent of a rule.

iii. Fuzzy Correlation (coined by us) measures the positive or negative relationship between the antecedent and the consequent and hence may be used to generate negative fuzzy rules based on an new threshold given by the user.

3.2 Applying Fuzzy Sets to Transaction Data

We also point out one of the extension from the work of Kuok et al (1998). As mentioned earlier, fuzzy set concept is applied to quantitative attribute data such as age, gender or salary. There has been no attempt to apply fuzzy concept to transaction datasets.

AR Mining literature typically generates rules from purchase transactions that do not mention quantities of items purchased. For example, buying bread and jam together and buying 3 loaves of bread and 1 bottle of jam, are quite different in interpretation.

In order to apply fuzzy concept to market basket, we need to convert the transaction into quantitative attributes. Consider the transaction set in Table 2 converted into quantitative attributes given in Table 3. We define a fuzzy set for the entire market transaction data as follows.

F = {High, Medium, Low}

Here High (H), Medium (M) and Low (L) represent quantities of items purchased by customers in each transaction. H is quantity 4-5 units, medium is 3 units and low is 1-2 units.

T.Id	Bread	Butter	Jam
T1	3	-	5
T2	4	2	3
T3	3	3	-

Table 2: Transaction set with quantities

T. Id	Bread			Butter			Jam		
	H	M	L	H	M	L	H	M	L
T1	0	1	0	0	0	0	1	0	0
T2	1	0	0	0	0	1	0	1	0
T3	0	1	0	0	1	0	0	0	0

Table 3: Quantitative representation of transaction data

T1: <Bread, Medium>, <Jam, High>

T2: <Bread, High>, <Butter, Low>, <Jam, Medium>

T3: <Bread, Medium>, <Butter, Medium>

We need to define a membership function for fuzzy sets H, M and L.

Let us consider the following fuzzy rule for which the membership votes are given in Table 4:

Bread, Medium → Jam, High

Bread, Medium	Jam, High
0.9	0.9
0.5	0.3
0.9	0

Table 4: Membership votes for antecedent and consequent of the rule

$$Significance = \frac{0.81 + 0.15 + 0}{3}$$
$$= 0.32$$

$$Certainty = \frac{0.32}{2.30}$$
$$= 0.14$$

$$Correlation = \frac{0.32 - (2.3)(1.2)}{(2.3)(1.2)}$$
$$= -0.88$$

This rule has 32% significance, 14% certainty and 88% negative correlation.

As a consequence negative rule Bread, Medium → ~ Jam, High gets generated.

The fuzzy approach opens a battery of issues.

i. It is not possible to have a uniform fuzzy set space across the entire transaction space. How does one handle it?

ii. How does one define the fuzzy operators apart from the typical fuzzy union,… thus leading to meaningful interestingness-based understanding of fuzzy ARs?

iii. Do the fuzzy ARs themselves need any further augmentation?

iv. How does one handle multiple fuzzy sets in a single transaction?

4. Business Applications of Negative ARs

AR Mining methods are useful across a variety of business applications. The data obtained from these business applications can contain both transactional and quantitative in nature. Interesting positive and negative rules can give important insights to managers and may enable in knowledge discovery about customer behaviour.

4.1 Market Basket Analysis

Market Basket Analysis (MBA) has the objective of individuating products, or groups of products, that tend to occur together in similar baskets (Giudici 2005). The data is mostly transactional in nature, representing baskets of each customer. The knowledge obtained from MBA may be used to reorganise a supermarket's layout for promotional campaigns and bundling of frequent products and new product development. MBA may also be used in e-commerce environments, where real-time modelling of an individual customer and personalized feedback is valuable (Apte, Liu, Pednault & Smyth 2002). Information from negative relationships among products may be used for clustering similar customers based on their purchase patterns. This leads to useful information that has potential to address managerial issues such as customer segmentation, personalization, forecasting and change detection. The more interesting the mined ARs, the more robust and accurate the solutions are ; thus promising significant economic payoff in the business world.

4.2 Customer Relationship Management

Swift (2001, p.12) define CRM as an ''enterprise approach to understanding and influencing customer behaviour through meaningful communications in order to improve customer acquisition, customer retention, customer loyalty, and customer profitability". Data mining techniques such as AR Mining, Classification and Prediction are used for extracting and identifying useful information and knowledge from enormous customer databases for making different CRM decisions (Berson et al., 2000). These customer databases usually comprise quantitative attributes such as demographics, age and loyalty card details. As such, the application of data mining techniques in CRM is worth pursuing in a customer-centric economy (Ngai et al 2011). Interesting positive and negative ARs can be used to build a model for predicting the value of a future customer (Wang et al., 2005). These rules can be applied to classify customers into loyal clients or those who abandon a company for competitors (Giudici 2005).

4.3 Credit Scoring

Credit scoring uses data mining techniques to evaluate the credit reliability of individuals who ask for credit when buying goods or services (Giudici 2005). Banks, Investment companies and Credit card organisations scan the customer database often comprising quantitative attributes to analyse customers' creditworthiness. The probability of loan repayment may be analysed using positive or negative ARs thus classifying creditors into two classes of risk: good and bad. This approach is similar to CRM where the past behaviour of an individual is scored in order to plan a future action.

References

1. Agrawal, R., & Srikant, R. (1994, September). Fast algorithms for mining association rules. In Proc. 20th int. conf. very large data bases, VLDB (Vol.1215, pp.487-499).

2. Agrawal, R., Imieliski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. ACM SIGMOD Record, 22(2), 207-216.

3. Antonie, M. L., & Zaïane, O. R. (2004). Mining positive and negative association rules: an approach for confined rules. In Knowledge Discovery in Databases: PKDD 2004 (pp.27-38). Springer Berlin Heidelberg.
4. Apte, C., Liu, B., Pednault, E. P., & Smyth, P. (2002). Business applications of data mining. *Communications of the ACM,* 45(8), 49-53.
5. Brin, S., Motwani, R., & Silverstein, C. (1997a, June). Beyond market baskets: Generalizing association rules to correlations. In ACM SIGMOD Record (Vol.26, No.2, pp.265-276). ACM.
6. Brin, S., Motwani, R., Ullman, J. D., & Tsur, S. (1997b, June). Dynamic itemset counting and implication rules for market basket data. In ACM SIGMOD Record (Vol.26, No.2, pp.255-264). ACM.
7. Dong, G., & Li, J. (1998). Interestingness of discovered association rules in terms of neighborhood-based unexpectedness. In Research and Development in Knowledge Discovery and Data Mining (pp.72-86). Springer Berlin Heidelberg.
8. Giudici, P. (2005). *Applied data mining: statistical methods for business and industry.* John Wiley & Sons.
9. Kuok, C. M., Fu, A., & Wong, M. H. (1998). Mining fuzzy association rules in databases. ACM Sigmod Record, 27(1), 41-46.
10. Mannila, H., Toivonen, H., & Verkamo, A. I. (1994, July). E cient algorithms for discovering association rules. In KDD-94: AAAI workshop on Knowledge Discovery in Databases (pp.181-192).
11. Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems,* 50(3), 559-569.
12. Padmanabhan, B., & Tuzhilin, A. (1999). Unexpectedness as a measure of interestingness in knowledge discovery. Decision Support Systems, 27(3), 303-318.
13. Savasere, A., Omiecinski, E., & Navathe, S. (1998, February). Mining for strong negative associations in a large database of customer transactions. In Data Engineering, 1998. Proceedings., 14th International Conference on (pp.494-502). IEEE.
14. Shaw, G., Xu, Y., & Geva, S. (2009). Interestingness Measures for Multi-Level Association Rules. Proceedings of ADCS 2009, 27-34.
15. Srikant, R., & Agrawal, R. (1995). Mining generalized association rules (pp.407-419). IBM Research Division.
16. Srikant, R., & Agrawal, R. (1996, June). Mining quantitative association rules in large relational tables. In Acm Sigmod Record (Vol.25, No.2, pp.1-12). ACM.
17. Yuan, X., Buckles, B. P., Yuan, Z., & Zhang, J. (2002). Mining negative association rules. In Computers and Communications, 2002. Proceedings. ISCC 2002. Seventh International Symposium on (pp.623-628). IEEE.