

A neural-networks associative classification method for association rule mining

P. Sermswatsri & C. Srisa-an

Faculty of Information Technology, Rangsit University, Thailand

Abstract

Association rule theory has been used to analyze relationships between data in a large database. It has been applied in various fields including market analysis, customer habit study, detection, classification, clustering and so on. However, it consumes a great amount of time. In the Apriori approach, all frequent patterns are found and then generate association rules from above frequent patterns. However, this approach generates a huge time-intensive query called “iceberg query”. Various researches have been done under the Apriori-like approach to improve performance of the frequent pattern mining task but the results were not as much as expected due to many scans on the dataset. This research aims to propose a flexible way of mining frequent patterns by extending the idea of the Associative Classification method. For better performance, the Neural-Network Associative Classification (NAC) method is proposed here to be one of the approaches for building accurate and efficient classifiers. In this paper, the Neural Networks Associative Classification method is used in order to improve its accuracy. The structure of the network reflects the knowledge uncovered in the previous discovery phase. The trained network is then used to classify unseen data. The accuracy rates obtained from the datasets show promising results.

Keywords: associative classification, association rule, data mining, back propagation neural networks.

1 Introduction

The current emphasis in Data Mining and Machine Learning research involves improving the classifier to be more precise and efficient. The important Machine Learning algorithms are Decision Trees [1], Rule Learning [2], Naïve-Bayes Classifier [3], and Statistics [4]. They use Heuristic Search and Greedy Search



techniques to find subsets of rules to create classifiers. This algorithm may make rules interesting, or not interesting, or may not be found. It reduces the classifiers accuracy. [5]

Then, Liu et al. [5] proposes a new algorithm called Associative Classification, which integrates two techniques, Data Classification [6] and Association Rules Discovery [7]. Selecting some of Class Association Rules: CARs from Association Rule to create a Classifier from [5] indicates that a Classifier from Association Rule is more exact than C4.5 [1] algorithm, but the algorithm proposed by Liu et al that is CBA (Classification Based on Associations) needs to have a ranking rule before it can create a Classifier. Ranking depends on the value of confidence and support of each rule. It makes the accuracy of CBA algorithm less precise than later algorithms such as CMAR (Classification based on Multiple Association Rules) and CPAR (Classification based on Predictive Association Rules).

2 Related theory

Applying Backpropagation Neural Network with Association Rules, it has the following related theory.

2.1 Association Classifications

Association Classifications integrates two major techniques in Data mining which are Data Classifications and Association Rule.

2.1.1 Data Classifications

Data Classifications is a kind of supervised learning algorithm by inputting some data to training and another one that is different from training to test the data. The result from training data is Data Classifier.

2.1.2 Association Rules

Determine:

$D = \{d_1, d_2, \dots, d_n\}$ is a database that consist of set of n data and $d \subseteq I$

$I = \{i_1, i_2, \dots, i_m\}$ is a set of all items that appear in D

This Association Rules has a format, which is $A \rightarrow B$ by support value = $s\%$ and confidence value = $c\%$ by $A, B \subseteq I$ and $A \cap B = \emptyset$ following conditions.

- Support Value is frequency of No. of data that consist of A and B or $P(A \cup B)$
- Confidence value is frequency of No. of data that consist of A and consist of B or $P(B|A)$
- Support value and Confidence value must show in percentage.

2.1.3 Class Associative Rules

Class Associative Rules is a rule of association from class data to identify association of an item set or attribute to a class of that data and the support value must be greater than the minimum support that is user defined.

Determine:



$D = \{d_1, d_2, \dots, d_i\}$ is a database that consist of set of data that have **n** attribute and

class label by $d = \{a_1, a_2, \dots, a_n, y_k\}; k=1, \dots, m$

$I = \{X_1, X_2, \dots, X_j\}$ is set of all items that appears in D

$Y = \{y_1, y_2, \dots, y_m\}$ is set of class label **m** class

Class Association that derive from Association Rule has a format, which is $X \rightarrow y$ by support value = $s\%$ and confidence value = $c\%$ by $X \subseteq I$ and $y \in Y$ following these condition

- Support value is the frequency of No. of all data that consist of item set X and Class label y
- Confidence value is the frequency of No. of data that consist of item set X and has Class label y also.
- Confidence value and Support value must show in percentage

2.1.4 Class Relation Rules following CBA algorithm

Class Relation Rules following CBA algorithm derive from Apriori algorithm [5] that is a basic algorithm to search Data Association Rules.

Determine:

- **Ruleitem** has following format $\langle (\text{condset}, \text{condsupCount}), (y, \text{rulesupCount}) \rangle$ that represents Association Rules $\text{condset} \rightarrow y$ by condset is item set and $y \in Y$
- **K-ruleitem** is ruleitem that **condset** consist of **k** members.
- **CondsupCount** is number of data that consist of **condset**.
- **RulesupCount** is number of data that consist of **condset** and has label type y .

Then

- Support value = $(\text{rulesupCount} / |D|) * 100\%$
- Confidence values = $(\text{rulesupCount} / \text{condsupCount}) * 100\%$

CBA algorithm that use to find Association Classifier can express in figure 1.

2.2 Neural Network

Neural Network [3] is one algorithm for machine learning. It has a concept from the human brain, the characteristic of a Neural Network is many neuron interconnect together where it makes many nodes and has weight in interconnection.

2.2.1 Backpropagation Neural Network

Backpropagation Neural Network is a type of multi-neuron. It uses many perceptron to connect together and use Backpropagation algorithm to adjust weight. Structure of Backpropagation:

Generally Backpropagation Neural Network use sigmoid activation function, then output of each node can be defined in this equation.

```

 $F_1 = \{\text{large 1-rule items}\};$ 
 $CAR_1 = \text{genRules}(F_1);$ 
 $prCAR_1 = \text{pruneRules}(CAR_1);$ 
for ( $k = 2; F_{k-1} \neq \emptyset; k++$ ) do
     $C_k = \text{candidateGen}(F_{k-1});$ 
    for each data case  $d \in D$  do
         $C_d = \text{ruleSubset}(C_k, d);$ 
        for each candidate  $c \in C_d$  do
             $c.\text{condsupCount}++;$ 
            if  $d.\text{class} = c.\text{class}$  then  $c.\text{rulesupCount}++;$ 
        end
    end
     $F_k = \{c \in C_k \mid c.\text{rulesupCount} \geq \text{minsup}\};$ 
     $CAR_k = \text{genRules}(F_k);$ 
     $prCAR_k = \text{pruneRules}(CAR_k);$ 
end
 $CARs = \cup_k CAR_k;$ 
 $prCARs = \cup_k prCAR_k;$ 

```

Figure 1: CBA algorithm that are used to find Association Classifier.

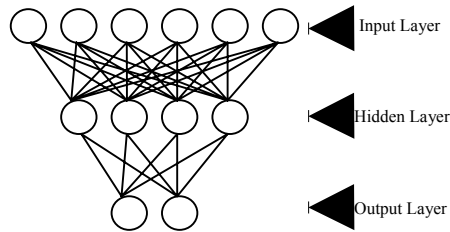


Figure 2: Structure of Backpropagation Neural Network.

$$\text{Output} = \sigma(\text{net}) = \frac{1}{1 + e^{-\text{net}}} \quad (1)$$

by

$$\text{net} = \sum_{i=0}^n (w_i x_i) \quad (2)$$

Backpropagation Neural Network learning is learning to adjust the weight to reduce training error, that is when a network gets data network and will calculate actual output to compare with the target output. If these weights give a bad result the network will feedback an error to input and recalculate output again by adjusting the weight and do it again and again until the result is near to the best solution or the error can be accepted. Weight adjusting in Backpropagation Neural Network algorithms are defined.

3 Applying a Backpropagation neural network to Associative Classifications

3.1 Class Associative Rules

This research adapts a CBA algorithm to find Association Classification without pruning rules that found, because Neural Network can learn efficiently even though training data have noises, expressed as Figure 3

An example of searching the Association Classification Rule following that algorithm can be expressed in Figure 3.

```

 $F_1 = \{\text{large 1-rule items}\};$ 
 $CAR_1 = \text{genRules}(F_1);$ 
for ( $k = 2; F_{k-1} \neq \emptyset; k++$ ) do
   $C_k = \text{candidateGen}(F_{k-1});$ 
  for each data case  $d \in D$  do
     $C_d = \text{ruleSubset}(C_k, d);$ 
    for each candidate  $c \in C_d$  do
       $c.\text{condsupCount}++;$ 
      if  $d.\text{class} = c.\text{class}$  then  $c.\text{rulesupCount}++$ 
    end
  end
   $F_k = \{c \in C_k \mid c.\text{rulesupCount} \geq \text{minsup}\};$ 
   $CAR_k = \text{genRules}(F_k);$ 
   $\text{prCAR}_k = \text{pruneRules}(CAR_k);$ 
end
 $CARs = \cup_k CAR_k;$ 

```

Figure 3: Algorithm to search Association Classification Rule.

Define data to search Association Classification Rule is detailed in Table 1.

Table 1: Example data to search Association Classification Rule.

	A	B	C
1.	p	s	Yes
2.	p	t	Yes
3.	q	s	Yes
4.	r	t	No
5.	q	t	No
6.	p	s	No
7.	p	s	Yes
8.	r	u	Yes
9.	r	u	No
10	q	u	No

- A and B are characteristic that A has possible values that are **p, q, r** and B has possible values are **s, t, u**
- C is Class label; possible values are **yes** and **no**.

When search Association Classification Rules by define minimum support = 15% and minimum confidence = 60%, we can express details as in Table 2.

Table 2: Searching Association classification Rules following proposed algorithm.

1st pass	F₁	$\langle \{(A,p)\}, 4 \rangle, \langle (C, \text{Yes}), 3 \rangle$
		$\langle \{(A,q)\}, 3 \rangle, \langle (C, \text{No}), 2 \rangle$
		$\langle \{(A,q)\}, 3 \rangle, \langle (C, \text{Yes}), 1 \rangle$
		$\langle \{(A,r)\}, 3 \rangle, \langle (C, \text{No}), 2 \rangle$
		$\langle \{(A,r)\}, 3 \rangle, \langle (C, \text{Yes}), 1 \rangle$
		$\langle \{(B,s)\}, 4 \rangle, \langle (C, \text{Yes}), 3 \rangle$
		$\langle \{(B,t)\}, 3 \rangle, \langle (C, \text{No}), 2 \rangle$
		$\langle \{(B,t)\}, 3 \rangle, \langle (C, \text{Yes}), 1 \rangle$
		$\langle \{(B,u)\}, 3 \rangle, \langle (C, \text{No}), 1 \rangle$
		$\langle \{(B,u)\}, 3 \rangle, \langle (C, \text{Yes}), 2 \rangle$
2nd pass	C₂	$\langle \{(A,p), (B,s)\}, (C, \text{Yes}) \rangle$
		$\langle \{(A,p), (B,t)\}, (C, \text{Yes}) \rangle$
		$\langle \{(A,q), (B,s)\}, (C, \text{Yes}) \rangle$
		$\langle \{(A,q), (B,t)\}, (C, \text{Yes}) \rangle$
		$\langle \{(A,q), (B,u)\}, (C, \text{No}) \rangle$
		$\langle \{(A,r), (B,t)\}, (C, \text{No}) \rangle$
	F₂	$\langle \{(A,p), (B,s)\}, 3 \rangle, \langle (C, \text{Yes}), 2 \rangle$
		$\langle \{(A,p), (B,t)\}, 1 \rangle, \langle (C, \text{Yes}), 1 \rangle$
		$\langle \{(A,q), (B,s)\}, 1 \rangle, \langle (C, \text{Yes}), 1 \rangle$
		$\langle \{(A,q), (B,t)\}, 1 \rangle, \langle (C, \text{Yes}), 1 \rangle$
		$\langle \{(A,q), (B,u)\}, 1 \rangle, \langle (C, \text{No}), 1 \rangle$
		$\langle \{(A,r), (B,t)\}, 1 \rangle, \langle (C, \text{No}), 1 \rangle$
CAR₁	$(A,p) \rightarrow (C, \text{Yes})$	
	$(A,q) \rightarrow (C, \text{No})$	
	$(A,r) \rightarrow (C, \text{No})$	
	$(B,s) \rightarrow (C, \text{Yes})$	
	$(B,t) \rightarrow (C, \text{No})$	
	$(B,u) \rightarrow (C, \text{No})$	
CAR₂		$\{(A,p), (B,s)\} \rightarrow (C, \text{Yes})$
CAR_s		CAR₁ U CAR₂

From table 2 we can express Association Classification Rules.

1. $(A,p) \rightarrow (C, \text{Yes})$ $s = 30\%$, $c = 75\%$
2. $(A,q) \rightarrow (C, \text{No})$ $s = 20\%$, $c = 66.7\%$

3. $(A,r) \rightarrow (C, \text{No})$ $s = 20\%$, $c = 66.7\%$
4. $(B,s) \rightarrow (C, \text{Yes})$ $s = 30\%$, $c = 75\%$
5. $(B,t) \rightarrow (C, \text{No})$ $s = 20\%$, $c = 66.7\%$
6. $(B,u) \rightarrow (C, \text{No})$ $s = 20\%$, $c = 66.7\%$
7. $\{(A,p), (B,s)\} \rightarrow (C, \text{Yes})$ $s = 20\%$, $c = 66.7\%$

When **s** and **c** are **support value** and **confidence value** respectively.

3.2 Create structure of Backpropagation Neural Network from Associative Classifications

Structure of Backpropagation Neural Network that input Association Classification to learning consists of 3 parts.

- Input layer is the first layer of a network that input Association Classification has to learn. In this layer it will input from each of the rules to be the input node by one characteristic and will be represented with one node, that is the input node with equal characteristics.
- Hidden layer is a neighbor layer with the input nodes connected together with “and”, that is each hidden node will be connected with the characteristic of each rule to collect each characteristic to be the complete rules, then the number of nodes in this layer are equal with the number of Association Classification to learning.
- Output layer is a layer that expresses the result of the network, each output node will be fully connected with every node in the hidden layer and the number of output node equal with the number of class of data that you need the network to learn.

3.3 Data creation to learning and adjust weight

Data creation will consider characteristic values that learns the consistency of characteristic value in each input and output node.

If characteristic of the data is identical with each characteristic, input and output node we will have the true value “True” represent with “1”, if that value conflicts with each characteristic input and output node we will have the true value “False” represented with “0”

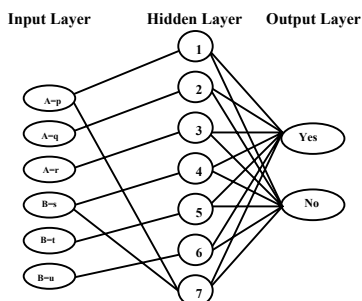


Figure 4: Structure of Backpropagation Neural Network from Association Classification Rules.

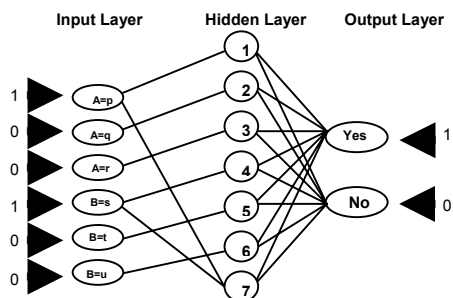


Figure 5: Data creation in Neural Network.

For example, data to learning is “A = p, B = s and has the class label “Yes” when inputting this data to a network will have the input value is “1 0 0 1 0 0” and output node is “1 0” as shown in Figure 5.

4 Experimental and result

Testing the proposed algorithm we will use data from **UCI Repository of Machine Learning Database** with 12 set to test. Classify data to learning data and testing data by 6-fold cross validation [3]. Details of the data are shown in Table 3.

Table 3: Data set to testing.

Data type	No. Data	Class	Characteristic
breast-w	699	2	10
cleve	303	2	13
diabetes	768	2	8
glass	214	7	10
heart	270	2	13
hepatitis	155	2	19
iris	150	3	4
lymph	148	4	18
pima	768	2	8
tic-tac	958	2	9
wine	178	3	13
zoo	101	7	17

Structure complexities of Neural Network in each data set depends on Association Classification Rules to learning in an experiment defined learning rate = 0.001 and momentum = 0.8 and learning will stop when the error s less than 0.05 or training 10,000 times.

In Figure 6 show compares result of each algorithm (native CBA, CPAR and proposed algorithm or NAC in the thick red line)

From the experiment above, NAC will get more efficient Association Classification Rules because Neural Network will learn to adjust weight to be input for the Association Classification Rules.

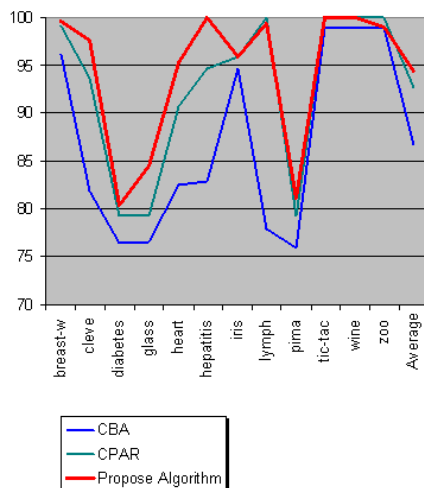


Figure 6: Compare the result of each algorithm and the proposed algorithm or NAC shown in thick red line, note that this line above almost matches the other algorithm.

5 Conclusion

Neural-Network Associative Classification (NAC) method is for building accurate and efficient classifiers in order to improve its accuracy. The structure of networks reflects the knowledge uncovered in the previous discovery phase. The trained network is then used to classify unseen data.

NAC that are presented in this research we get Data Classifier from Class Association that more precisely match than the old one by learning rules and adjusting weight of Class Association to find the best Class Association and use that one to classify data.

In our future work, we intend to find optimize minimal support in order to pruning.

References

- [1] J. R. Quinlan, *C4.5: Programs for Machine Learning*: Morgan Kaufmann, 1993.
- [2] P. Clark and T. Niblett, "The CN2 induction algorithm", *Machine Learning*, 3, pp. 261–283, 1989.



- [3] T. M. Mitchell, *Machine Learning*: The McGraw-Hill Companies Inc., 1997.
- [4] T. S. Lim, W. Y. Loh and Y. S. Shih, "A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms", *Machine Learning*, 40, pp. 203-228, 2000.
- [5] B. Liu, W. Hsu, and Y. Ma, "Integrating Classification and Association Rule Mining", *KDD'98*, pp. 80-86, 1998.
- [6] J. Han and M. Kamber, *Data Mining Concepts and Techniques*: Morgan Kaufmann, 2000.
- [7] R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules between Sets of Items in Large Databases", *SIGMOD'93*, pp. 207-216, 1993.
- [8] Napaparn Yingchankul and Yachai Limpiyakorn, *Applying Backpropagation Neural Networks to Associative Classification*: Chulalongkorn University, 2000.