

# Joining Associative Classifier for Medical Images

Jiang Yun<sup>1,2</sup>, Li Zhanhuai<sup>1</sup>, Wang Yong<sup>1</sup>, and Zhang Longbo<sup>1</sup>

<sup>1</sup> College of Computer Science, Northwestern Polytechnical University,  
710072, Xi'an, P.R.China  
jiangyun@mail.nwpu.edu.cn lizhh@nwpu.edu.cn

wangyong@mail.nwpu.edu.cn zhanglb@sdut.edu.cn

<sup>2</sup> College of Mathematics and Information Science, Northwest Normal University,  
730070, Lanzhou, P.R.China

## Abstract

*One of the best prevention measures against breast cancer is the early tumor detection in digital mammography. Detecting tumor in mammography is a difficult task because of their size and the high content of similar patterns in the image. This brings the necessity of creating automatic tools to find whether a mammography present tumor or not. In this paper we join association rule classifier with rough set theory which we call the joining associative classifier (JAC) to mining digital mammography. The experimental results shows that this joining associative classifier performance at 77.48% of classifying accuracy which is higher than 69.11% using associative classifier only. At the same time, the number of rules decreased distinctively. Moreover, the experiments we conducted demonstrate the use and effectiveness of association rule mining in image categorization.*

## 1. Introduction \*

Breast cancer is the second cause of death in women in the world. This is why breast cancer is considered a public health problem. Currently there are no methods to prevent breast cancer, which is why early detection represents a very important factor in cancer treatment and allows reaching a high survival rate. Mammography is considered the most reliable method in early detection of breast cancer. Due to the high volume of mammograms to be read by physicians, the accuracy rate tends to decrease, and automatic reading

of digital mammograms becomes highly desirable. That is why the computer aided diagnosis systems are necessary to assist the medical staff to achieve high efficiency and effectiveness [1].

A lot of efforts have been done to solve this problem. In [1], the authors use data mining techniques to detect and classify anomalies in the breast. They use neural networks and association rules as the data mining algorithms. They also use a histogram equalization process to enhance the image contrast, and then they perform a feature extraction process and include those features combined with patients' information for the classification task. Their results were of 82.248% with neural network and 69.11% with association rules. But one noticeable advantage of the association rule-based classifier is the time required for training, which is very low compared to other methods such as neural networks [1].

In [2], the authors use a Bayesian network to find and classify regions of interest. In this research they use a segmentation algorithm based on wavelet transforms and the use of a threshold that corresponding to the local minimum of the image. In [3], the authors present a method for feature extraction of lesions in mammograms using an edge based segmentation algorithm. In [4], we find an evaluation of different methods that can be used to get texture features from regions of interest extracted from mammogram images. In [5], we can see how the wavelet transformation has been used to detect groups of micro-calcifications in digital mammograms. In addition, some other methods were presented in the literature based on fuzzy set theory [6] and Markov models [7] and neural networks [8]. With all this effort, there is still no widely used method to classify

---

\* This paper is supported by national science foundation No. 60373108

medical images. This is because this medical domain requires high accuracy and low training time.

Association rule mining is one of the most important tasks in data mining and it has been extensively studied and applied to market basket analysis. In [1], the authors have presented association rule mining method for classify anomalies in the breast. But their classify accuracy was only 69.11%. In this paper, we present a new classification method for medical images based on association rule mining algorithm—Apriori and rough set theory which we call the joining associative classifier (JAC). It is tested on real datasets MIAS[9] (the Mammographic Image Analysis Society) and get 77.48% classify accuracy which is higher than 69.11% using Apriori algorithm only in [1].

The rest of the paper is organized as follows: Section 2 describes the images pre-processing phase and the feature extraction phase. Section 3 presents the new association rule-based classifier— joining associative classifier (JAC). In section 4, we present our experiments and results. Finally, in section 5, we show our conclusions and future work.

## 2.Data Pre-processing and Feature Extraction

This section summarizes the techniques used to enhance the mammograms as well as the features that were extracted from images. To compare with [1], we use almost same experimental methods as it.

### 2.1. Data Pre-processing

Pre-processing is always a necessity whenever the data to be mined in noisy, inconsistent or incomplete. Pre-processing significantly improves the effectiveness of data mining techniques [10]. The type size of the images in MIAS was 1024x1024 and almost 50% of the whole image comprised the background with a lot of noise. In addition, these images are scanned at different illumination conditions, and therefore some images appeared too bright and some were too dark. The first step toward noise removal was pruning the images with a cropping operation. The second step was an image enhancement. Thus, we eliminated almost all the background information and most of the noise. An example of cropping that eliminates the artefacts and the black background is given in Figure 1 (a-b)

Since the resulting images had different sizes, the  $x$  and the  $y$  coordinates were normalized to a value between 0 and 255. The cropping operation was done automatically by sweeping horizontally through the image. Then we applied the Histogram Equalization method to enhance the image in order to diminish the

effect of over-brightness or over-darkness in images. Histogram Equalization increases the contrast range in an image by increasing the dynamic range of grey levels [10]. Figure 1 (c) shows an example of histogram equalization result after cropping.

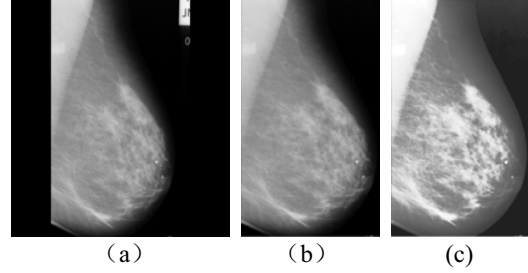


Fig 1. Pre-processing phase on an example image: (a) original image; (b) cropped image; (c) enhanced image based on histogram equalization

### 2.2. Feature Extraction

After pre-processing the images, features relevant to the classification are extracted from the cleaned images. The extracted features are organized in a database, which is the input for the mining phase of the classifier. This database is also constructed by merging some already existing features like the type of the tissue (dense, fatty and fatty-glandular) and the location of the abnormality (like the center of a circle surrounding the tumor). The extracted features are four statistical parameters: *mean*, *variance*, *skewness* and *kurtosis*. The formula for the statistical parameters computed is the following:

$$\text{Mean: } \mu = \sum_{k=1}^N f_k p_f(f_k) \quad (1)$$

$$\text{Variance: } \sigma^2 = \sum_{k=1}^N (f_k - \mu)^2 p_f(f_k) \quad (2)$$

$$\text{Skewness: } \mu_3 = \frac{1}{\sigma^3} \sum_{k=1}^N (f_k - \mu)^3 p_f(f_k) \quad (3)$$

$$\text{Kurtosis: } \mu_4 = \frac{1}{\sigma^4} \sum_{k=1}^N (f_k - \mu)^4 p_f(f_k) \quad (4)$$

Where  $N$  denotes the number of gray levels in the mammogram,  $f_k$  the  $k$ th gray level and  $p_f(f_k) = n_k / n$ , where  $n_k$  is the number of pixels with  $f_k$  gray level and  $n$  is the total number of pixels in the region.

All these extracted features are computed over smaller windows of the original image. The original image is first split in four parts. For a more accurate extraction of the features we split each of these four

regions in other four parts. The statistical parameters were computed for each of the sixteen sub-parts of the original image [1]. After that, we get sixty-four statistical features.

### 3. Joining Associative Classifier

In this section we propose to execute behind to the Apriori algorithm the Rough Set Theory in order to discover strong dependencies that can, in general, be useful to reduce the original set of attributes and obtain more powerful rules to classify the concept. The following are the basic concepts of the rough set theory and association rules.

#### 3.1. Rough Set Theory

The original Rough Set Theory was proposed by Pawlak [13][14]. This theory is concerned with analysis of deterministic data dependencies.

**Definition 1. Information Systems :** In the Rough Set Theory, information systems are used to represent knowledge. An *information system*  $S = (U, A, V, f)$  consists of :

$U$  — a nonempty, finite set named *universe*, which is a set of objects,  $U = \{x_1, x_2, \dots, x_m\}$ ;

$A$  — a nonempty, finite set of *attributes*,  $A = C \cup D$ , in which  $C$  is the set of *condition attributes*, and  $D$  is the set of *decision attributes*;

$V = \bigcup_{a \in A} V_a$ ,  $V$  is the *domain* of  $a$ ;

$f : U \times A \rightarrow V$  — an *information function*. For each  $a \in A$  and  $x \in U$ , an information function  $f(x, a) \in V_a$  is defined, which means that for each object  $x$  in  $U$ ,  $f$  specify its attribute value.

**Definition 2. Lower and Upper Approximation:** Due to imprecision which exists in the real world data, there are always conflicting objects contained in a decision table. Here *conflicting objects* refers to two or more objects that are undistinguishable by employing any set of condition attributes, but they belong to different decision classes. Such objects are called *inconsistent*. Such a decision table is called *inconsistent decision table*. In the rough set theory, the approximations of sets are introduced to deal with inconsistency. If  $S = (U, A, V, f)$  is a decision table, suppose  $B \subseteq A$ , and  $X \subseteq U$ , then the *B-lower* and *B-upper approximations* of  $X$  are defined as:

$$\underline{B(X)} = \bigcup \{Y \in U / IND(B) : Y \subseteq X\},$$

$$\overline{B(X)} = \bigcup \{Y \in U / IND(B) : Y \cap X \neq \emptyset\}.$$

Here,  $U/IND(B)$  denotes the family of all equivalence classes of  $B$ ;  $IND(B) = \{(x, y) \in U \times U \mid \forall a \in B, f(x,$

$a) = f(y, a)\}$  is the *B-indiscernibility relation*.  $\underline{B(X)}$  is the set of all elements of  $U$  which can be certainly classified as elements of  $X$ , employing the set of attributes  $B$ . The *Positive Region* of  $X$  is defined as:

$$POS_B(X) = \underline{B(X)} \quad (5)$$

$\overline{B(X)}$  is the set of elements of  $U$  which can be possibly classified as elements of  $X$  using the set of attributes  $B$ . The set  $Bnd_B(X) = \overline{B(X)} - \underline{B(X)}$  is called the *B-boundary* of  $X$ . If  $Bnd_B(X) = \emptyset$ , then we say that  $X$  is definable on  $B$ ; otherwise we say that  $X$  is non-definable on  $B$ , which is also named as *rough set* [13][14].

#### 3.2. Association Rules

The purpose of association discovery is to find items that imply the presence of other items. An association rule is formally described as follows [11]:

Let  $I = \{i_1, i_2, \dots, i_n\}$  be a set of literals called items. Let  $D$  be a set of transactions, each transaction  $T \subset I$ . An association rule is an implication of the form  $X \rightarrow Y$  where  $X \subset I$  and  $Y \subset I$  and  $X \cap Y = \emptyset$ . The rule  $X \rightarrow Y$  holds in the transaction set  $D$  with confidence  $c$  if  $c\%$  of transactions in  $D$  that contain  $X$  also contain  $Y$ . The rule  $X \rightarrow Y$  holds in the transaction set  $D$  with support  $s$  if  $s\%$  of transactions in  $D$  contain  $X \cup Y$ . Give a set of transaction  $D$  the problem of mining association rules is to generate all association rules that have support and confidence greater than the user-specified minimum support (*min-sup*) and minimum confidence (*min-conf*). In order to derive the association rules two steps are required: 1) Find the large itemsets for a given *min-sup*; 2) Compute rules for a given *min-conf* based on the itemsets obtained before [12]. In this paper, we are based on the Apriori algorithm and rough set theory to obtain the association rules.

#### 3.3. The Algorithm of Joining Associative Classifier

**Algorithm:** Joining Associative Classifier— JAC

**Input:** —  $S$  the input decision data table. The set of attributes  $A$  will be divided in condition attribute set  $C$  and decision attribute set  $D$ .  
— *min-conf* the minimum confidence for rules.  
—  $n$  maximum number of condition allowed in the antecedent of the rules.

**Output:** — Data mining rule set  $D_{Str}$ .

We will assume that input table  $S$  has been discretized and binarized. The process that will be performed is as follows:

Step1. Execute the Apriori algorithm on the input data table  $S$ . Internally the algorithm will calculate the size of large itemsets  $k$  depending on the nature of  $S$ . We call this new set of rules  $S_{ass}$ .

Step2. Delete all those rules in  $S_{ass}$  which the decision attribute occurs as part of their antecedent.

Step3. If  $\forall i \exists k$  such that  $A_k \rightarrow D_i$  then  $A_k$  is superfluous attribute so that it can be removed from  $A$ . Remove also such rules from  $S_{ass}$ .

Step4. Analyse of the rules in the new  $S_{ass}$ . This set of rules contains two kinds of rules that we have called *strong classification rules* and *meta-rules*. The former is composed by all those rules in  $S_{ass}$  which the consequent is some  $D_i$  with confidence  $\geq \text{min-conf}$ . The latter is a set containing rules that will allow us to reduce the number of condition attributes.

- Include *strong classification rules* in the output data mining rule set  $D_{Str}$ .
- Those association rules in *meta-rules* that only contain condition attributes have to be taken into account as they highlight dependencies among condition attributes. We will call this set of rules  $S_{red}$ .

Step5. Reduce  $A$  taking into account the rules in  $S_{red}$ .

Step6. Execute the *Positive Region* algorithm to obtain a set of rules that will be included in output data mining rule set  $D_{str}$ .

## 4. Experimental Results

This section introduces the data collection that we used and the experimental results obtained using the joining associative classifier proposed.

### 4.1. Mammography Collection

The data collection that was used in our experiments was taken from the Mammographic Image Analysis Society (MIAS)[9]. This same collection has been used in [1]. It consists of 322 images, which belong to three big categories: normal, benign and malign. There are 208 normal images, 63 benign and 51 malign, which are considered abnormal. In addition, the abnormal cases are further divided in six categories: microcalcification, circumscribed masses, speculated

masses, ill-defined masses, architectural distortion and asymmetry. All the images also include the locations of any abnormalities that may be present. The existing data in the collection consists of the location of the abnormality (like the center of a circle surrounding the tumor), its radius, breast position (right or left), type of breast tissues (fatty, fatty-glandular and dense) and tumor type if it exists (benign or malign). All the mammograms are medio-lateral oblique view. We selected this dataset because it is freely available, and to be able to compare our method with other published work like [1], since it is a commonly used database for mammography categorization.

### 4.2 Experimental Results

We used the 10 fold cross validation techniques to evaluate the algorithm performance. We divided the features database in ten splits. For each split we selected about 90% of the dataset for training and the rest for testing. That is 288 images in the training set and 34 images in the testing set. The features database is composed with the extracted features and the existing data of 322 images in MIAS. All the numeric attributes are discrete using algorithm DBChi2 [15]. In the training phase, the joining associative classifier was applied on the training data and the set of association rules were extracted. Then, for an image in the testing set, the classification process searches in **Table 1**. The comparison of two algorithm on MIAS

Ten splits	Apriori		JAC	
	rules	accuracy	rules	accuracy
1	965	63.705	66	69.342
2	1080	67.648	98	86.373
3	1005	80.412	71	77.586
4	847	64.705	85	72.912
5	1782	64.705	74	78.224
6	1201	67.642	70	77.055
7	1359	88.205	69	77.691
8	895	64.721	80	73.752
9	1383	66.512	73	82.123
10	1366	62.901	76	79.819
average	1188	69.11%	76	77.48%

this set of rules for finding the class that is the closest to be attached with the object presented for categorization.

The support was set to 10% and the confidence to 3%. The reason for choosing the 3% percent for the confidence is motivated by the fact that the database has more normal cases (about 70%). As we can see in table 1, we got 77.48% classify accuracy result with the joining associative classifier than 69.11% with the association rule alone [1]. At the same time, the number of rules decreased distinctively.

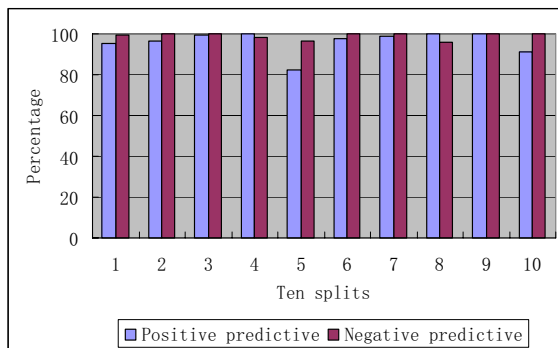
In table 1, the first column is the ten splits of dataset. The second and third columns are the number of rules and classify accuracy on Apriori. Then , the forth and fifth columns are the number of rules and classify accuracy on JAC.

We also present the positive predictive and the negative predictive graphs in Figure 2 to show that both false positive and false negative are very small for normal cases, which means that for abnormal images we have a very small number of false negative which is very desirable in medical classification. The formula for the positive predictive and the negative predictive are given below:

$$\text{Positive predictive} = \text{TP}/(\text{TP} + \text{FP}) \quad (6)$$

$$\text{Negative predictive} = \text{TN}/(\text{TN} + \text{FN}) \quad (7)$$

Where TP stands for true positives, FP for false positives, FN for false negatives and TN for true negatives. In addition, we can notice from equations 6 and 7 that the values for FP and FN tend to zero when the positive predictive and the negative predictive tend to 100%. Thus, from the Figure 2, we can see that the positives predictive and in particular the negative predictive are almost 100% with our approach which means that the classification in just normal and abnormal categories was actually high.



**Fig 2.** Positive and negative predictive over the ten splits

The experiment is made on a computer with single 996MHz Pentium 3 CPU and 256MB memory, and the JAC program is written in Borland C++ Builder 6. The source program of Apriori is downloaded from [16].

## 5. Conclusions

In this paper, we presented a cooperative association classify method applied to medical image classification. This joining associative classifier is based on the association rule mining algorithm---Apriori and the rough set theory. This approach provides a sound basis for the definition of a new cooperative algorithm to obtain comprehensible rules,

while avoiding the computational complexity of classical predictive methods. In addition, we demonstrated how important the image pre-processing phase is in building a classifier. The evaluation of the classifier was carried out on MIAS dataset and experimental results show that the accuracy of the joining associative classifier reaches 77.48% than 69.11% which execute Apriori algorithm itself in [1]. And the positive predictive value and the negative predictive value tend to 100% in more than half the splits. There are some future research directions to be studied. To cooperate with medical staff would get more interesting results. In addition, the extraction of different features or a different database organization could lead to improved results.

## 6. References

- [1] Maria-Luiza Antonie, Osmar R.Zaiane and Alexandru Coman . *Application of data mining techniques for medical image classification*. In Proc. of Second Intl. Workshop on Multimedia Data Mining (MDM/KDD'2001) in conjunction with Seventh ACM SIGKDD, San Francisco, USA, 94-101, 2001.
- [2] Zhang, Xiao-Ping and Desai, Mita D. *Wavelet Based Automatic Thresholding for Image Segmentation*. In Proceedings of the ICIP'97 conference, Santa Barbara, CA, 26-29, 1997.
- [3] Bottigli, U. and Golosio, B. *Feature Extraction from Mammographic Images Using Fast Matching Methods*. Nuclear Instruments and Methods in Physics Research. ELSEVIER, A 487, 209-215 , 2002.
- [4] Sharma, M and Singh, S. *Evaluation of Texture Methods for Image Analysis*. Proceedings of the 7<sup>th</sup> Australian and New Zeland Intelligent Information Systems Conference. Perth, November, 18-21, 2001.
- [5] Yoshida, H.Doï, K.Nishikawa, R.Muto, K. and Tsuda, M. *Application of the Wavelet Transform to Automated Detection of Clustered Microcalcifications in Digital Mammograms*. Academic Reports of Tokyo Institute of Polytechnics, 16, 24-37, 1994.
- [6] D.Brazokovic and M.Neskovic. *Mammogram Screening Using Multiresolution-based Image Segmentation*. International Journal of Pattern Recognition and Artificial Intelligence, 7(6):1437-1460, 1993.
- [7] H.Li et al. *Markov Random Field for Tumor Detection in Digital Mammography*. IEEE Trans. Medical Imaging, 14(3):565-576, 2000.
- [8] I.Christoyianni et al. *Fast Detection of Masses in Computer-aided Mammography*. IEEE Signal Processing Magazine, 54-64, 2000.
- [9]<http://www.wiau.man.ac.uk/services/MIAS/MIASweb.html>
- [10] Rafael C.Gonzalez and Richard. E. Woods. *Digital Image Processing*. Addison-Wesley, 1993. second edition.

- [11] Jiawei Han and Michline Kamber. *Data mining concepts and Techniques*. Morgan Kaufmann, 2001.
- [12] Agrawal R., Imielinski T., Swami A. *Mining Association Rules between sets of Item in Large Database*, Proceedings of ACM SIGMOD, 207-216, 1993.
- [13] Pawlak Z.W. *Rough sets and intelligent data analysis*. Information sciences, 1-12, 2002.
- [14] Pawlak Z.W. *Rough classification*. Int.J.Human-Computer Studies, 369-383, 1999.
- [15] Hu X., and Cercone N.. *Data Mining Via Generalization , Discretization and Rough Set Feature Selection* [J]. Knowledge and Information System: An International Journal, 1999, 1(1).
- [16] Zhi-Hua Zhou. AI Softwares&Codes (Maintained by Zhi-Hua Zhou) .  
[http://cs.nju.edu.cn/people/zhoush/zhoush.files/ai\\_resouce/software.htm](http://cs.nju.edu.cn/people/zhoush/zhoush.files/ai_resouce/software.htm), 2004-02-29