Expert Systems with Applications 36 (2009) 3465–3469

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

An expert system for detection of breast cancer based on association rules and neural network

Murat Karabatak [a,*], M. Cevdet Ince [b]

[a] Fırat University, Department of Electronics and Computer Science, 23119 Elazig, Turkey
[b] Fırat University, Department of Electric-Electronics Engineering, 23119 Elazig, Turkey

ARTICLE INFO

Keywords:
Association rules
Neural network
Automatic detection
Breast cancer

ABSTRACT

This paper presents an automatic diagnosis system for detecting breast cancer based on association rules (AR) and neural network (NN). In this study, AR is used for reducing the dimension of breast cancer database and NN is used for intelligent classification. The proposed AR + NN system performance is compared with NN model. The dimension of input feature space is reduced from nine to four by using AR. In test stage, 3-fold cross validation method was applied to the Wisconsin breast cancer database to evaluate the proposed system performances. The correct classification rate of proposed system is 95.6%. This research demonstrated that the AR can be used for reducing the dimension of feature space and proposed AR + NN model can be used to obtain fast automatic diagnostic systems for other diseases.

1. Introduction

Data classification process using knowledge obtained from known historical data has been one of the most intensively studied subjects in statistics, decision science and computer science. It has been applied in problems of medicine, social science management and engineering. Variable problems such as disease diagnosis, image recognition, and credit evaluation using classification techniques (Michie, Spiegelhalter, & Tayor, 1994). In medical and other domains, linear programming approaches were efficient and effective methods (Bennett & Mangasarian, 1992; Freed & Glover, 1981; Grinold, 1972; Smith, 1968). Recently, intelligent methods such as NN and support vector machines have been intensively used for classification tasks (Ryua, Chandrasekaranb, & Jacobc, 2007).

One of the application areas of analyzing database and pattern recognition is automated diagnostic systems. The aims of these studies are assisting to doctors in making diagnostic decision. Thanks to modern facilities, very large databases can be collect in medicine. These databases need special techniques for analyzing, processing and effective use of them. Data mining and knowledge discovery in database are an approach to find relationships buried in data (Choua, Leeb, Shaoc, & Chenb, 2004). The methodologies consist of data visualization, machine learning and statistical techniques and these can be summarized as classification, prediction, clustering, etc. (Curt, 1995).

Breast cancer is a very common and serious cancer for women. Mammography is one of the most used methods to detect the breast cancer (Choua et al., 2004). In literature, radiologists show considerable variation in interpreting a mammography (Elmore et al., 1994). Fine needle aspiration cytology (FNAC) is also widely adopted in the diagnosis of breast cancer. But, the average correct identification rate of FNAC is only 90%. So, it is necessary to develop better identification method to recognize the breast cancer. Statistical techniques and artificial intelligence techniques have been used to predict the breast cancer by several researchers (Kovalerchuck, Triantaphyllou, Ruiz, & Clayton, 1997; Pendharkar, Rodger, Yaverbaum, Herman, & Benner, 1999). The objective of these identification techniques is to assign patients to either a benign group that does not have breast cancer or a 'malignant' group who has strong evidence of having breast cancer. So, breast cancer diagnostic problems are more general and widely discussed classification problem. (Anderson, 1984; Dillon & Goldstein, 1984; Hand, 1981; Johnson & Wichern, 2002).

There are many techniques to predict and classification breast cancer pattern. In Choua et al. (2004), artificial neural network and multivariate adaptive regression splines approach was used to classify the breast cancer pattern. In Aragonés, Ruiz, Jiménez, Pérez, and Conejo (2003), a combined neural network and decision trees model was used for prognosis of breast cancer relapse. In Ryua et al. (2007), isotonic separation technique was used to predict breast cancer. In Şahan, Polat, Kodaz, and Güneş (2007), a new hybrid method based on fuzzy-artificial immune system and k-nn algorithm was proposed for breast cancer diagnosis. And in Übeyli (2007), Wisconsin breast cancer data was classified using multilayer perceptron neural network, combined neural network, probabilistic neural network, recurrent neural network and support vector machine.

* Corresponding author.
E-mail addresses: mkarabatak@firat.edu.tr (M. Karabatak), mcince@firat.edu.tr (M.C. Ince).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.02.064

In this study, an AR + NN method was proposed to use in breast cancer diagnosis problem. This method consists of two-stages. In the first stage, the input feature vector dimension is reduced by using association rules. This provides elimination of unnecessary data. In the second stage, neural network uses these inputs and classifies the breast cancer data.

2. Wisconsin breast cancer database

Breast cancer is the most common cancer among women; excluding non melanoma skin cancers. This cancer affects one in eight women during their lives. It occurs in both men and women, although male breast cancer is rare. Breast cancer is a malignant tumor that has developed from cells of the breast. Although scientists know some of the risk factors (i.e. ageing, genetic risk factors, family history, menstrual periods, not having children, obesity) that increase a woman's chance of developing breast cancer, they do not yet know what causes most breast cancers or exactly how some of these risk factors cause cells to become cancerous. Research is under way to learn more and scientists are making great progress in understanding how certain changes in DNA can cause normal breast cells to become cancerous (Übeyli, 2007).

In this study, the Wisconsin breast cancer database was used and analyzed. They have been collected by Dr. William H. Wolberg (1989–1991) at the University of Wisconsin–Madison Hospitals. There are 699 records in this database. Each record in the database has nine attributes. The nine attributes detailed in Table 1 are graded on an interval scale from a normal state of 1–10, with 10 being the most abnormal state. In this database, 241 (65.5%) records are malignant and 458 (34.5%) records are benign.

3. Preliminaries

3.1. Association rules

In order to see how AR can be used in breast cancer data with NN, first of all it is needed to define AR. AR find interesting associations and/or relationships among large set of data items. AR show attributes value conditions that occur frequently together in a given dataset. They allow capturing all possible rules that explain the presence of some attributes according to the presence of other attributes. A typical and widely-used example of association rule mining is Market Basket Analysis (Agrawal et al., 1993).

Let $I = (i_1, i_2, \ldots, i_m)$ be a set of literals, called items. Let D be a database of transaction, where each transaction T is a set of items such that $T \subseteq I$. For a given itemset $X \subseteq I$ and a given transaction T, we say that T contains X if and only if $X \subseteq I$. The support count of an itemset X is defined to be $\sup_x$ = the number of transactions in D that contain X. we say that an itemset X is large, with respect to a support threshold of $s\%$, if $\sup_x \geqslant |D| \times s\%$, where $|D|$ is the number of transactions in the database D. An association rule is an implication of the form "$X \Rightarrow Y$", where $X \subseteq I$, $Y \subseteq I$ and $X \cap Y = \emptyset$. The AR "$X \Rightarrow Y$" is said to hold in database D with confidence $c\%$ if no less than $c\%$ of the transactions in D that contain X also contain Y. The rule $X \Rightarrow Y$ has support $s\%$ in D if $\sup_{X \cup Y} = |D| \times s\%$ (Pendharkar et al., 1999). Thus, AR aims at discovering the patterns of co-occurrence of attributes in a database. For instance, an association rule in a supermarket basket data may be in 10% of transactions, 85% of the people buying milk also buy yoghurt in that transaction. AR may be useful in many applications such as business applications, market basket analysis, store layout and promotion on the items, telecommunication alarm correlation, university course enrollment, texture and image processing (Karabatak, Şengür, İnce, & ve Türkoğlu, 2006).

Table 1
Wisconsin breast cancer data description of attributes

Attribute number	Attribute description	Values of attributes	Mean	Standard deviation
1	Clump thickness	1–10	4.42	2.82
2	Uniformity of cell size	1–10	3.13	3.05
3	Uniformity of cell shape	1–10	3.20	2.97
4	Marginal adhesion	1–10	2.80	2.86
5	Single epithelial cell size	1–10	3.21	2.21
6	Bare nuclei	1–10	3.46	3.64
7	Bland chromatin	1–10	3.43	2.44
8	Normal nucleoli	1–10	2.87	3.05
9	Mitoses	1–10	1.59	1.71

N = 699 observations, 241 malignant and 458 benign.

3.1.1. Apriori algorithm

The Apriori algorithm is a state of the art algorithm most of the association rule algorithms are somewhat variations of this algorithm (Agrawal et al., 1993). The Apriori algorithm works iteratively. It first finds the set of large 1-item sets, and then set of 2-itemsets, and so on. The number of scan over the transaction database is as many as the length of the maximal item set. Apriori is based on the following fact: The simple but powerful observation leads to the generation of a smaller candidate set using the set of large item sets found in the previous iteration. The Apriori algorithm presented in Agrawal and Srikant (1994) is given as follows:

```
Apriori()
L1 ={large 1-itemsets}
k = 2
while Lk−1≠ϕ do
begin
    Ck = apriori_gen(Lk−1)
    for all transactions t in D do
    begin
        Ct = subset(Ck, t)
        for all candiate c ∈ Ct do
            c.count = c.count+1
    end
    Lk = {c ∈ Ck|c.count ⩾ minsup}
    k = k + 1
end
```

Apriori first scans the transaction databases D in order to count the support of each item i in I, and determines the set of large 1-itemsets. Then, iteration is performed for each of the computation of the set of 2-itemsets, 3-itemsets, and so on. The kth iteration consists of two steps (Rushing, Ranganath, Hinke, & Graves, 2002):

- Generate the candidate set C_k from the set of large (k-1)-itemsets, L_{k-1}.
- Scan the database in order to compute the support of each candidate itemset in C_k.

The candidate generation algorithm is given as follows:

```
Apriori_gen (Lk−1)
Ck = ϕ
for all itemsets X ∈ Lk−1 and Y ∈ Lk−1 do
    if X1 = Y1 ∧ ··· ∧ Xk−2 = Yk−2 ∧ Xk−1 < Yk−1 then begin
        C = X1X2 ... Xk−1Yk−1
        add C to Ck
end
delete candidate itemsets in Ck whose any subset is not in LK−1
```

The candidate generation procedure computes the set of potentially large k-itemsets from the set of large (k-1)-itemsets. A new candidate k-itemset is generated from two large (k-1)-itemsets if their first (k-2) items are the same. The candidate set C_k is a superset of the large k-itemsets. The candidate set is guaranteed to include all possible large k-itemsets because of the fact that all subsets of a large itemset are also large. Since all large itemsets in L_{k-1} are checked for contribution to candidate itemset, the candidate set C_k is certainly a superset of large k-itemsets. After the candidates are generated, their counts must be computed in order to determine which of them are large. This counting step is really important in the efficiency of the algorithm, because the set of the candidate itemsets may be possibly large. Apriori handles this problem by employing a hash tree for storing the candidate. The candidate generation algorithm is used to find the candidate itemsets contained in a transaction using this hash tree structure. For each transaction T in the transaction database D, the candidates contained in T are found using the hash tree, and then their counts are incremented. After examining all transaction in D, the ones that are large are inserted into L_k (Karabatak et al., 2006).

3.2. Neural networks

Neural networks (NNs) are biologically inspired and mimic the human brain. They are occurring neurons. These neurons are connected each other with connection links. These links have weights. They multiplied with transmitted signal in network. The output of each neuron is determined by using an activation function such as sigmoid and step. Usually nonlinear activation functions are used. NN's are trained by experience, when applied an unknown input to the network it can generalize from past experiences and product a new result (Bishop, 1996; Hanbay, Turkoglu, & Demir, 2007; Haykin, 1994). The output of the neuron net is determined by Eq. (1). A simple artificial neuron model is shown in Fig. 1.

$$y(t+1) = a\left(\sum_{j=1}^{m} w_{ij}x_j(t) - \theta_i\right) \quad \text{and} \quad f_i \triangleq \text{net}_i = \sum_{j=1}^{m} w_{ij}x_j - \theta_i \qquad (1)$$

where $X = (X_1, X_2 \ldots X_m)$ represent the m input applied to the neuron, W_i represent the weights for input X_i, θ_i is a bias value, $a(\cdot)$ is activation function. NNs models have been used for pattern matching, nonlinear system modeling, communications, electrical and electronics industry, energy production, chemical industry, medical applications, data mining and control because of their parallel processing capabilities. When designing a NN model a number of considerations must be taken into account. First of all the suitable structure of the NN model must be chosen, after this the activation function, the number of layers and the number of units in each layer must be chosen. Generally desired model consist of a number of layers. The most general model assumes complete interconnections between all units. These connections can be bidirectional or unidirectional.

4. Applications

Feature extraction is the key for pattern recognition so that it is arguably the most important component of designing the intelligent system based on pattern recognition since even the best classifier will perform poorly if the features are not chosen well. A feature extractor should reduce the pattern vector (i.e., the original waveform) to a lower dimension, which contains most of the useful information from the original vector (Türkoğlu, Arslan, & Ilkay, 2003). Fig. 2 shows the proposed automatic detection system block diagram. It consists of two parts: (a) feature extraction and redcution with AR (b) classification with NN.

4.1. AR Layer

AR is a method to find the associations and/or relationships among items in large databases. So, we can use it to detect relations among inputs of any system and later eliminate some unnecessary inputs. We propose two different techniques to eliminate inputs. These are named as AR1 and AR2, respectively.

4.2. AR1

The AR1 technique uses all input parameters and their all records to find relations among the input parameters. If we find rules that have enough support value and high confidence value, than we can eliminate some inputs thanks to these rules. In the AR form ($X \Rightarrow Y$), Y itemset also depend on X itemset. Thus, we can eliminate all items in Y itemset. So, these are not necessary to use in NN inputs.

4.3. AR2

Especially, we can use AR2 with classification problems. AR2 uses all input parameters but not all their records. We find only large itemsets for every class. All items in these large itemsets are most important items to classification. Thus, we can only use these items to classify all data. If an item of large itemset of any class is large in other classes and it has different value, this item must be used as NN inputs.

In this study, we used AR1 and AR2 to reduce the number of NN inputs for breast cancer detection problem. We eliminated only one input parameter of NN By using AR1 technique. Because, one of the rules is

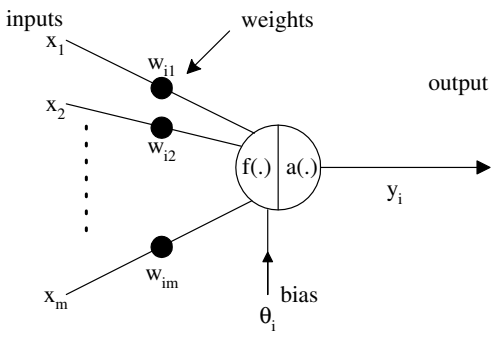

Fig. 1. Artificial neuron model.

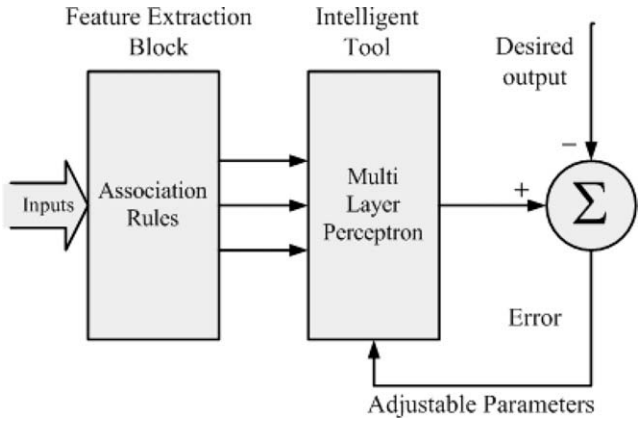

Fig. 2. The block diagram of the automatic detection system.

Table 2
MLP architecture and training parameters

Architecture	
The number of layers	3
The number of neuron on the layers	Input: 4, 8, 9 Hidden: 11 Output: 1
The initial weights and biases	Random
Activation functions	Tangent-sigmoid Tangent-sigmoid Linear
Training parameters	
Learning rule	Levenberg–Marquardt Back-propagation
Sum-squared error	0.01

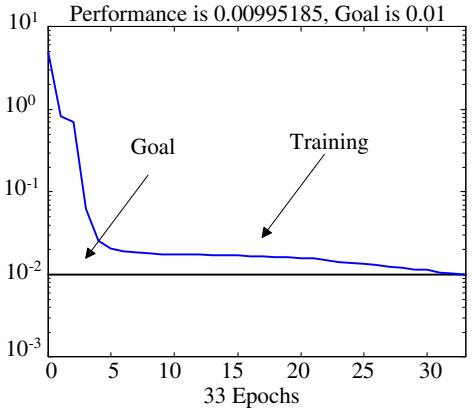

Fig. 3. Neural Network training performance.

Input: 1-3-8-9⇒2
Value: 1-1-1-1⇒1 confidence is 100%.

According to this rule; if the value of 1st, 3rd, 8th and 9th input parameters are 1, the value of 2nd input parameter is 1. Then it says that 2nd input already depend on others. So we did not use 2nd input parameter in NN input.

Wisconsin breast cancer database has two classes. These are benign and malignant classes. Using AR2, we found large itemsets of benign and malignant classes given as follows:

Input: 2-8-9
Value: 1-1-1 (large itemsets for benign class)
Input: 6
Value: 10 (large item for malignant class)

According to this large itemset, we can say that 2nd, 8th and 9th input parameters already can define benign class and 6th input parameter can define malignant class. These parameters are the most important parameters for breast cancer detection problems. So, we only used these inputs in NN.

4.4. NN Layer

Multi-layer perceptron (MLP): the intelligent classification is realized in this layer by using features, which are obtained from AR layer. The training parameters and the structure of the MLP used in this study are shown in Table 2. These were selected for the best performance, after several experiments. Fig. 3 shows the AR2 + NN training performance.

Table 3
Performance comparison for breast cancer detection using NN, AR1 + NN and AR2 + NN

The classifier	The epochs	Correct classified	Miss classified	Correct classification rate (%)
NN (9, 11, 1)	61	216	11	95.2
AR1 + NN (8, 11, 1)	44	221	6	97.4
AR2 + NN (4, 11, 1)	33[a]	217	10	95.6

[a] Goal is 0.01.

5. Modeling results

This study was performed using Wisconsin breast cancer database with 9 attributes and 699 records. In test stage, 3-fold cross validation method was applied and average values were calculated. The performance comparison and correct classification rates are tabulated in Table 3.

As shown in Table 3, the best classification performance was obtained with AR1 + NN with eight inputs and its correct classification rate is 97.4%. The correct classification rate of NN with 9 inputs is 95.2% and the correct classification rate of AR2+NN is 95.6% was obtained. So, we can use AR1+NN for best classification performance and AR2 + NN for using input parameters at minimum number.

6. Conclusions

In this study, an automatic diagnosis system for detecting breast cancer based on association rules (AR) and neural network (NN) is presented. Feature extraction is the key for pattern recognition and classification. The best classifier will perform poorly if the features are not chosen well. A feature extractor should reduce the feature vector to a lower dimension, which contains most of the useful information from the original vector. So, AR is used for reducing the dimension of breast cancer database and NN is used for intelligent classification. The proposed AR + NN system performance is compared with NN model. The dimension of input feature space is reduced from nine to four by using AR. In test stage, 3-fold cross validation method was applied to the Wisconsin breast cancer database to evaluate the proposed system performances. The correct classification rate of proposed system is 95.6% for four inputs and 97.4% for eight inputs. This research demonstrated that the AR can be used for reducing the dimension of feature vector and proposed AR + NN model can be used to obtain efficient automatic diagnostic systems for other diseases.

References

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th international conference on very large databases* (pp. 487–499).
Agrawal, R., Imielinski T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the ACM SIGMOD international conference on management of data*.
Anderson, T. W. (1984). *An introduction to multivariate statistical analysis*. New York: Wiley.
Aragonés, M. J., Ruiz, A. G., Jiménez, R., Pérez, M., & Conejo, E. A. (2003). A combined neural network and decision trees model for prognosis of breast cancer relapse. *Artificial Intelligence in Medicine, 27*, 45–63.
Bennett, K. P., & Mangasarian, O. L. (1992). Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software, 1*, 23–34.
Bishop, C. M. (1996). *Neural networks for pattern recognition*. Oxford: Clarendon Press.

Choua, S.-M., Leeb, T.-S., Shaoc, Y. E., & Chenb, I.-F. (2004). Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications, 27*, 133–142.

Curt, H. (1995). The devil's in the detail: Techniques: Tools, and applications for database mining and knowledge discovery-Part. *Intelligent Software Strategies*, 1–15.

Dillon, W. R., & Goldstein, M. (1984). *Multivariate analysis methods and applications*. New York: Wiley.

Elmore, J., Wells, M., Carol, M., Lee, H., Howard, D., & Feinstein, A. (1994). Variability in radiologists interpretation of mammograms. *New England Journal of Medicine, 331*(22), 1493–1499.

Freed, E., & Glover, F. (1981). A linear programming approach to the discriminant problem. *Decision Sciences, 12*(1), 68–74.

Grinold, R. C. (1972). Mathematical programming methods of pattern classification. *Management Science, 19*(3), 272–289.

Hanbay, D., Turkoglu, İ., & Demir, Y. (2007). An expert system based on wavelet decomposition and neural network for modeling Chua's circuit. *Expert Systems with Applications*. doi:10.1016/j.eswa.2007.03.002.

Hand, D. J. (1981). *Discrimination and classification*. New York: Wiley.

Haykin, S. (1994). *Neural networks, a comprehensive foundation*. New York: Macmillan College Publishing Company Inc.

Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.

Karabatak, M., Şengür, A., İnce, M. C., & ve Türkoğlu, İ. (2006). *Association rules for texture classification*. IMS.

Kovalerchuck, B., Triantaphyllou, E., Ruiz, J. F., & Clayton, J. (1997). Fuzzy logic in computer-aided breast-cancer diagnosis: Analysis of lobulation. *Artificial Intelligence in Medicine, 11*, 75–85.

Michie, D., Spiegelhalter, D. J., & Tayor, C. C. (1994). *Machine learning, neural and statistical classification*. London: Ellis Horwood.

Pendharkar, P. C., Rodger, J. A., Yaverbaum, G. J., Herman, N., & Benner, M. (1999). Associations statistical, mathematical and neural approaches for mining breast cancer patterns. *Expert Systems with Applications, 17*, 223–232.

Rushing, J. A., Ranganath, H. S., Hinke, T. H., & Graves, S. J. (2002). Image segmentation using association rule features. *IEEE Transactions on Image Processing, 11*, 558–566.

Ryua, Y. U., Chandrasekaranb, R., & Jacobc, V. S. (2007). Breast cancer prediction using the isotonic separation technique. *European Journal of Operational Research, 181*, 842–854.

Şahan, S., Polat, K., Kodaz, H., & Güneş, S. (2007). A new hybrid method based on fuzzy-artificial immune system and k-nn algorithm for breast cancer diagnosis. *Computers in Biology and Medicine, 37*, 415–423.

Smith, F. W. (1968). Pattern classifier design by linear programming. *IEEE Transactions on Computers C-17*(4), 367–372.

Türkoğlu, İ., Arslan, A., & Ilkay, E. (2003). An intelligent system for diagnosis of the heart valve diseases with wavelet packet neural networks. *Computers in Biology and Medicine, 33*, 319–331.

Übeyli, E. D. (2007). Implementing automated diagnostic systems for breast cancer detection. *Expert Systems with Applications, 33*, 1054–1062.