

A VISION-BASED INSPECTION SYSTEM USING FUZZY ROUGH NEURAL NETWORK METHOD

MENG-XIN LI^{1,2}, CHENG-DONG WU², FENG JIN³

¹Faculty of Information and Control Engineering Shenyang Jianzhu University, Shenyang 110168, China

²Northeastern University, Shenyang 110004, China

³Shenyang Huitian Redian Co.Ltd. 110004

E-MAIL: limengxin1972@yahoo.com.cn

Abstract:

A vision-based inspection method based on rough set theory, fuzzy set and neural network algorithm is presented. The rough set method is proposed to remove redundant features for its data analysis and processing. The reduced data is fuzzified to represent the feature data in a more suitable form for input to a BP network classifier. The BP neural classifier is considered the most popular, effective and easy-to-learn model for complex, multi-layered network. By the experiment research, the hybrid method shows good classification accuracy and short running time, which are better than the results using BP network and neural network with fuzzy input.

Keywords:

Vision-based inspection; classification; fuzzy input; rough set; neural Network

1. Introduction

In the wood mill, the veneer sheets are placed on a conveyor, which runs at a speed of 2.2m/s and the sheets appear at one to two second intervals for human inspection. This task is extremely stressful and a little disturbance or loss of attention will result in a misclassification. Huber et al [1] made a series of experiments and found an accuracy of 68% with human inspection of boards. Similar experiments carried out by Polzleitner and Schwingshagl [2] indicated an accuracy of 55%. It is imperative to develop an automatic visual inspection system to relieve the human inspector and improve the classification accuracy, thus improving the productivity and profitability of the plywood factory.

Various pattern recognition methods have been researched during the last four decades, such as statistical decision theory, syntax and neural networks. Hybrid methods are of great interest due to their proven adaptability and advantages. This research presents a method using fuzzy rough neural network method. The

rough set method is proposed to remove redundant features for its data analysis and processing. The reduced data is fuzzified to represent the feature data in a more suitable form for input to a BP network classifier. The BP neural classifier is considered the most popular, effective and easy-to-learn model for complex, multi-layered network.

2. Rough set for feature selection of wood veneer defect

2.1. Basic concepts

The following terms or concepts are introduced in order to facilitate and understand the proposed algorithm.

Information System: It is assumed that the given set of training samples represents the knowledge about the domain. In the approach described here, the training set is described by a classification system. The objects in a universe U are described by a set of attribute values.

Formally, an information system S is a quadruple $\langle U, A, V, f \rangle$, where $U = \{x_1, x_2, \dots, x_N\}$ is a finite set of objects, which in this case are states of the environment; A is a finite set of attributes and the attributes in A are further classified into two disjoint subsets, condition attributes C and decision attributes D , such that $A = C \cup D$ and $C \cap D = \emptyset$; $V = \bigcup_{a \in A} V_a$ is a set of attribute values and V_a is the domain of attribute a (the set of values of attribute a); and $f: U \times A \rightarrow V$ is an information function which assigns particular values from domains of attributes to objects such that $f(x_i, a) \in V_a$ for all $x_i \in U$ and $a \in A$.

Decision Table: An information system can be designed as a decision table if the attribute set is divided into two disjoint sets - condition attribute set C and decision attribute set D , and $C, D \subset A$.

Indiscernibility Relation: For every set of attributes $B \subset A$, an indiscernibility relation $IND(B)$ is defined in the

following way: two objects, x_i and x_j , are indiscernible by the set of attributes B in A , if $b(x_i)=b(x_j)$ for every $b \in B$. The equivalence class of $IND(B)$ is called the elementary set in B because it represents the smallest discernible group of objects. For any element x of U , the equivalence class of x_i in relation $IND(B)$ is represented as $[x_i]_{IND(B)}$. The equivalent class of relation $IND(C)$ and relation $IND(D)$ are called condition class and decision class respectively for condition attribute set C and decision attribute set D .

Lower and Upper Approximation: Let X denote a subset of elements of the universe U ($X \subset U$). The lower approximation of X in B ($B \subseteq A$), denoted as \underline{BX} , is defined as the union of all these elementary sets which are contained in X . More formally:

$$\underline{BX} = \{x_i \in U \mid [x_i]_{IND(B)} \subset X \neq \emptyset\} \quad (1)$$

The above statement is to be read as: the lower approximation of the set X is a set of objects x_i , which belongs to the elementary sets contained in X . The upper approximation of the set X , denoted as \overline{BX} , is the union of these elementary sets, which have a non-empty intersection with X :

$$\overline{BX} = \{x_i \in U \mid [x_i]_{IND(B)} \cap X \neq \emptyset\} \quad (2)$$

For any object x_i of the lower approximation of X , it is certain that it belongs to X . For any object x_i of the upper approximation of X , it may belong to X .

$BNX = \overline{BX} - \underline{BX}$ is called a boundary of X in U .

CORE: The set of all attributes indispensable in C is denoted by $CORE(C)$.

$$CORE(C) = \bigcap RED(C) \quad (3)$$

where $RED(C)$ is the set of all reducts of C .

Accuracy of Approximation: An accuracy measure of the set X in $B \subseteq A$ is defined as

$$\mu_B(X) = card(\underline{BX}) / card(\overline{BX}) \quad (4)$$

The cardinality of a set is the number of objects contained in the lower (upper) approximation of the set X , $0 \leq \mu_B(X) \leq 1$.

2.2. A rough sets feature selection method

For an information system for wood veneer classification, there are 17 condition attributes including mean grey level, mode grey level, median grey level, standard deviation, skewness, kurtosis, lower number of pixels, higher number of pixels, lower grey level, higher grey level, dark grey level, bright grey level, number of edge pixels (threshold= μ), number of edge pixels (threshold= $\mu-2\delta$), number of edge pixel for feature 14, number of edge pixels (threshold= $\mu+2\delta$), number of edge pixel for feature 16, and 1 decision attribute expressed by 1-13 that means 13 defects including holes, pin knots, rotten

knots, roughness, splits, strips, discoloration, coloured strips, barks, worms holes, curly grain, clear wood and sound knots. The task is to find out the optimal attributes and acquire decision rules.

2.2.1. Data discretization

It is necessary to process the attribute values with a discretization algorithm to express and simplify the decision table. Clustering is a useful tool for analysing the structure of attribute spaces, and deducing similarity and dissimilarity among the observations. In terms of its high dimensionality, discovering clusters of arbitrary shapes, and dealing with different types of attributes, a hierarchical clustering method is presented [3] and adopted for data discretization before attribute reduction. This method has advantages such as embedded flexibility regarding the level of granularity and ease of handling any forms of similarity or distances. It can be divided into agglomerative and divisive algorithms. The agglomerative algorithm usually produces a sequence of clustering schemes of a decreasing number of clusters at each step, which results from merging the two closest clusters. The agglomeration schedule can be visualised by a dendrogram that shows which samples are combined. Nevertheless, it is not certain how many clusters are in the data for a group can be merged into different clusters. So the number of clusters further comes from the idea about statistics such as PST2, PSF, CCC which can judge how many classes should be suitable. Chosen clusters should make the number of clusters as small as it could be. All the optimal values constitute the information system.

The 17 defect features of wood veneers, used as condition attributes $C=\{X_1, X_2 \dots X_{17}\}$, are discretized using hierarchical clustering and the values of decision attributes are expressed by 1-13.

Table 1. The representation of discretized knowledge

U	b	c	d	e	f	g	h	...	r	Y
1	2	1	2	1	1	1	3	...	1	1
2	1	1	2	1	2	2	1	...	1	1
3	2	1	1	1	2	2	2	...	1	1
4	1	1	2	1	1	1	1	...	1	1
5	2	1	2	1	2	2	3	...	2	1
6	1	1	2	1	2	2	1	...	1	1
7	1	1	1	1	1	2	1	...	2	1
8	1	1	1	1	2	2	1	...	2	1
9	1	1	1	1	3	1	1	...	1	1
...
232	2	2	1	1	1	1	2	...	1	13

2.2.2. Attribute reduction

The aim of attribute reduction is to find a minimal subset of related attributes that preserves the classification accuracy of the original attributes of the decision table. It is therefore necessary to identify important attributes. There are many reducts, but in most cases it is not necessary to find all the reducts. The reduct with the least number of combinations of its attributes is selected [4]. In this research, 17 condition attributes are reduced according to the consistency principle in the following steps.

Step 1: The repetitive samples in the decision table are merged.

Step 2: The data in the decision table is further processed through attribute reduction based on the consistency principle. Important attribute sets thus remain.

The principle of rough sets for reducing redundant attributes can be expressed as the following:

Supposing $C=\{X_1, X_2, \dots, X_n\}$ is an attribute set, if $\text{POS}_C(D) = \text{POS}_{(C-\{X_i\})}(D)$, then X_i in C is omissible or superfluous; otherwise attribute X_i in C is indispensable.

This has been implemented, as shown in Fig. 1. The reduct of condition attributes determines whether there are different decision values when the attribute values are the same in the decision table. Based on consistency principle, if an attribute set is removed and harmony still remains unchanged, the attribute set is removable. Among the 17 attributes of wood veneers, {c}, {g}, {i}, {l}, {m}, {q} are omissible after the attribute reduction, and the remaining attributes are sufficient. Mode grey level {c} and Kutosis {g} are insensitive to defining the changes of pixels. Because only splits and holes belong to white defects among the 13 defects, {i}, {m} and {q} are considered omissible. Dark grey level {l} contributes little to classification decision. After the attribute reduction, the optimum attributes are combined to obtain the decision rules for classification.

3. A neural network with fuzzy input for wood veneer inspection

3.1. Fuzzifier

A fuzzy set [4] can be represented as membership function μ_A that associates with each element x of the universe of discourse X , a number $\mu_A(x)$, i.e. membership grade, in the interval $[0, 1]$. In particular, $\mu_A: A \rightarrow [0, 1]$, where set A can also be treated as a subset of X . The main function of the fuzzifier maps a crisp input point $x \in X$ into a fuzzified value in $A \in U$ (the universe). There are two types of fuzzifier:

- Singleton fuzzifier: fuzzy set A with support x_i , where $\mu_A(x_i)=1$ for $x=x_i$ and $\mu_A(x_i)=0$ for $x \neq x_i$, for which the input measurement x is perfect crisp, $i=1,2,\dots,n$.
- Non-singleton fuzzifier: $\mu_A(x_i)$ reaches maximum value 1 at $x=x_i$ and decreases from 1 to 0 while moving away from $x=x_i$.

The determination of the fuzzy membership function is the most important issue in applying a fuzzy approach. No common approach is available for determining such a function. In some cases, the fuzzy membership function is attained subjectively as a model.

3.2. A Neural network algorithm with fuzzy input

Taking advantage of data processing using fuzzy sets, fuzzy theory is considered to combine with a BP algorithm. Suppose that we have a recognition problem of m classes, which has m nodes in the output layer. The weighted distance is first defined between the i th class and the j th class:

$$z_{ji} = \sqrt{\sum_{k=1}^n \left(\frac{x_{jk} - m_{ik}}{\sigma_{ik}} \right)^2}, \quad i = 1, 2, \dots, m \quad (5)$$

where x_{jk} is the k th vector of the j th pattern vector; $1/\sigma_{ik}$ is a normalisation factor, which results in small class weights for high variance, and m_i and σ_i are the mean value and standard dispersion respectively. The ambiguity of the j th pattern belonging to the i th class is then defined as follows:

$$\mu_{ij} = \frac{1}{1 + (Z_{ji} / \alpha)^\beta}, \quad i = 1, 2, \dots, m \quad (6)$$

where α and β are parameters used for controlling the fuzzy degree and $\alpha, \beta > 0$. According to equation 6.5, there is a low attributive degree if there is a large distance between a pattern and a class. If all elements satisfy $\mu_{ij} \neq 0$, a high fuzziness exists. If only one element satisfies $\mu_{ij} \neq 0$, no fuzziness exists. Under the condition of a high fuzziness, there is a need to modify the ambiguity factor in order to enlarge the difference of membership function.

$$\mu_{ij,INT} = \begin{cases} 2(\mu_{ij})^2 & 0 \leq \mu_{ij} \leq 0.5 \\ 1 - 2(1 - \mu_{ij})^2 & \text{others} \end{cases} \quad (7)$$

For the j th pattern, x_j , the i th subvector of the desired output, y_j is defined as:

$$y_{ij,INT} = \begin{cases} \mu_{ij,INT} & \text{high fuzziness} \\ \mu_{ij} & \text{others} \end{cases} \quad (8)$$

where $0 \leq y_{ij} \leq 1$. All the input and desired output vectors (x_j, y_j) can be used for training with the improved BP neural network.

4. A classifier using rough sets based neural network with fuzzy input

The input data is dealt with using rough sets and fuzzy sets for their powerful function of disposing infinite and

incomplete information. This will decrease the number of input nodes and complexity of neural network. A classifier using rough sets based neural network with fuzzy input is proposed (Figure 1).

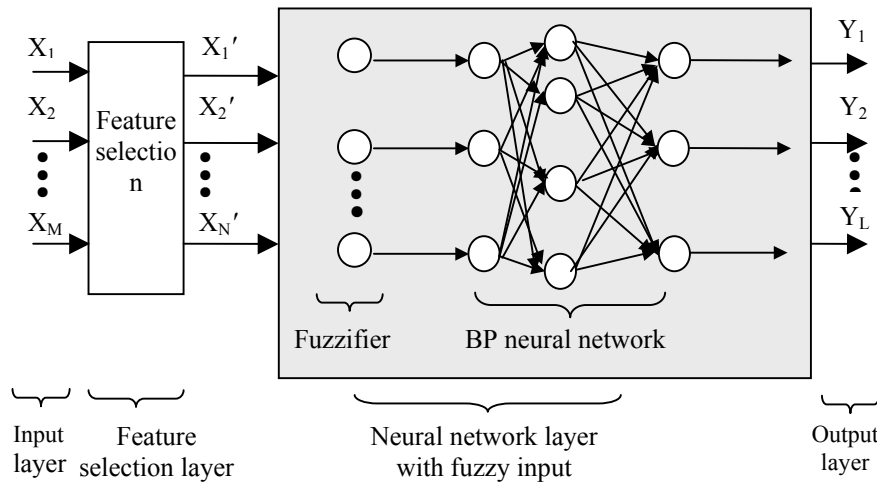


Figure 1. A classifier using rough sets based neural network with fuzzy input

The system includes input layer, data reduction layer, neural network layer with fuzzy input and output layer.

- Input data

This raw data is preprocessed to suppress noise and normalise the input, and the processed data is input to the feature reduction layer. The normalisation formula is

$$Z = \frac{x - \mu}{3\delta} \quad (9)$$

where x is raw data, μ is the sample mean, δ is the sample deviation and Z is the normalised data which is restricted to $[-1, 1]$ [Kjell et al., 1995].

- Feature reduction layer

The normalised data is processed with a hierarchical clustering discretization method. The data reduction is performed using rough sets, and important features for classification remain in the data.

- Neural network layer with fuzzy input

This layer includes a fuzzifier for the reduced data and the improved BP neural network. The crisp input data is converted into fuzzy data through fuzzification, and the fuzzified data is input to the improved BP neural network for defect classification. It is expected that a high classification accuracy and rapid running speed will be achieved through fast data processing.

- Output layer

Output results are obtained from this layer. The maximum coding method is used for the classification decision, which sets the highest output value to 1 and the others to 0. In other words, the defect class corresponds to the output neuron with the largest value.

4.1. Experiments

Sample data from wood veneer defects is used for testing, of which 80% is used as training data and 20% as testing data. The experiments are carried out in the 3 groups of data. The results of feature reduction are used by the neural network classifier with fuzzy input.

The fuzzified data is then input to the BP network for training and testing. It should be noted it is not necessary to change the learning rate by large amounts. Especially, the closer the training to convergence is, the smaller the learning rate change should be. Therefore a learning coefficient that is very close to 1, but less than 1, 0.999, is chosen for adjusting the weight slowly. The learning rate η is expressed below:

$$\eta(k+1) = 0.999\eta(k) \quad (10)$$

where k is the number of epochs and maximum of k is 6000, and initial learning rate is set to 0.5. The experimental result is shown in Figure 2, which takes Group 2 of sample data as an example.

For comparison, the BP method with fuzzy input and BP network are adopted and trained with the original sample data, and comparison results are provided in Table2.

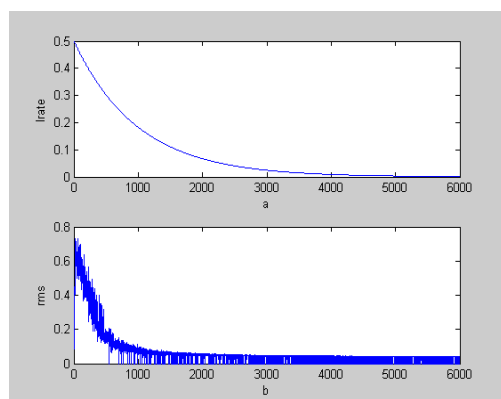


Figure 2. Training results of the RNNFI classifier with group 2

Table 2 comparison of overall accuracy and running time

	Accuracy	Average times	Average epochs
BP network	86.01%	324.7s	44323
BP network with fuzzy input	89.12%	14.79s	5773
A fuzzy rough neural network	96.97%	9.67s	4322

The results are further improved with each method proposed in that the average accuracy of classification increases from 86.01, 89.13% to 96.97% and the average running time drops from 324.7, 14.79s to 9.67s. Obviously, the rough fuzzy neural network method has the best classification performance.

5. Conclusions

To summarise, the classifier has several obvious characteristics which listed below:

- The neural network acquires knowledge through the process of training. The BP algorithm is a mature and effective one..
- An important idea of rough sets is to remove redundant attributes where there is a large amount of sample data. Elimination of redundant attributes can help identify strong, non-redundant classification rules. The rough set method for feature selection is particularly useful for dealing with imprecise, inconsistent and overlapping information.
- Fuzzy set theory is an expressive tool for coding. It entails a low computational load compared to neural

networks and makes decisions in a simple and robust way. Data fuzzination applied to the neural network classifier can also reduce the system running time while improving the classification accuracy.

- The hybrid method incorporating a neural network, fuzzy sets and rough sets can achieve a high classification accuracy and rapid speed taking advantage of the complementary characteristics of these techniques even in situations where data is imprecise, noisy, inconsistent and huge. There is no need for extra hardware to deal with uncertainties.

This paper has presented the rough sets based neural network with fuzzy input for pattern recognition as a more effective hybrid approach. In the feature selection process, redundant features are reduced significantly without losing essential information using rough set. In the feature classification, data fuzzification is used to deal with imprecise data and shorten the running time, and the improved BP neural network tackles the local minimum problem to achieve a good accuracy. The hybrid method has taken the advantages of all the techniques incorporated.

Experimental results have shown that the rough fuzzy neural network classifier has a high classification accuracy of wood veneer defects and a short running time. The method is considered general and can be applied to inspection of other products such as ceramic tiles.

References

- [1] Huber H.A., Mcmilin C.W. and Mckinney J.P., "Lumber defect detection abilities of furniture rough mill employees", *Forest Products Journal*, 35 (11/12), pp.79-82. 1985.
- [2] Polzleitner W. and Schwingshagl G., "Real-time surface grading of profiled wooden boards", *Industrial Metrology*, (2), pp.283-298.,1992.
- [3] Li M.X., Wu C.D. and Yue, Y. "A hierarchical clustering method for attribute discretization in rough set theory". *Proceedings of International Conference on Machine Learning and Cybernetics (ICMLC)*, Shanghai, China, 26-29 August, Vol.6, pp.3650-3654, 2004.
- [4] Zadeh L.A., "Fuzzy sets, Information and Control", Vol.8, pp.338-352, 1965.
- [5] Kjell B., Woods W.A. and Freider O., "Information retrieval using letter tuples with neural network and nearest neighbour classifiers", *IEEE International Conference on Systems, Man and Cybernetics*, Vancouver, Canada, pp.1222-1226, 1995.