Batik Image Classification Rule Extraction using Fuzzy Decision Tree

Anita Ahmad Kasim*, Retantyo Wardoyo**
* Department of Electronic Engineering, Tadulako University
** Department of Computer Science and Electronics Instrumentation, Faculty of Mathematic and Natural Science, Gadjah Mada University

Keywords:

Classification rule
Fuzzy Decision Tree
Batik
Pattern Recognition

ABSTRACT

Batik is an Indonesian cultural heritage that recognized as an international cultural heritage. Batik as one of the world's cultural heritage can be preserve through the Batik Image Pattern Recognition. One of the things that important in pattern recognition is classification rule. In the pattern recognition necessary, clear determination of classification rules to be able to recognize a pattern well. Classification of data needed to identify potentially characteristics in the set N objects contained in a database. It is categorized them into groups different. This study aims to obtain the classification rules by using fuzzy decision tree. This study begins by acquiring image data batik (pre-processing), enhancement and segmentation (process), feature extraction include entropy, energy, contras and homogeneity. Next, calculate the value of information gain and construct a decision tree and finally extracting fuzzy classification rules batik patterns. Extract classification rules that can classify patterns of batik and batik pattern instead.

Corresponding Author:

First Author,
Doctoral Student of Department of Computer Science and Electronics, Faculty of Mathematic and Natural Science, Universitas Gadjah Mada,
Email: anita.ahmad.k@ugm.ac.id cc: nita.kasim@gmail.com

1. INTRODUCTION

Decision tree classifiers work with data whose values are known and precise. Classifiers are handle data with uncertain information. Value uncertainty arises in many applications during the data collection process [1]. Decision tree is an approach used in machine learning. As the demands for Classification of the images is getting higher and higher, users want to search the images more fast and precisely, so a kind of fuzzy decision tree algorithm was presented[2]. The decision tree has been proving as one tool that can do the description, classification and generalization of data. It is relate to the density and clarity representation of a classification function. Many methods were develops with the aim of constructing a decision tree from a collection of data. The data set obtained sometimes ambiguous, vague, therefore, it is possible to combine these two methods to obtain the classification rules is more valid.

Classification is a function of the learning process that maps the data into a few classes that have been define previously. Each class classification based on inductive learning algorithm is give a set of inputs in the form of examples consisting of attributes and values corresponding class [3]. Classification is one method of data mining to predict the class label in a record in the data. The use of decision tree method allows the prediction of an object has more than one class. The fuzzy decision tree has employed many studies [4] [5]. Fuzzy decision tree research was also conducted in addition, the research value of the data to classify students also been carried out [3]. In this research, the classification of the batik pattern using fuzzy decision tree. It will obtain the extract classification rules that can classify the patterns of batik and batik pattern instead.

Fuzzy set was first introduced by Zadeh to represent and manipulate data that have the non-statistical uncertainty. Special fuzzy set theory to discuss the calculation and reasoning using natural language has ambiguous meanings. Fuzzy sets based on the idea to expand the range of characteristic functions such that the function will include real numbers on the interval [0 1]. Membership value indicates

that an item in the universe of discourse is not only to be at 0 or 1, but also the value that lies in between. The truth value of an item is not only a true or false value where a value of 0 indicates false and a value of 1 indicates the truth value true. There are values that lie between right and wrong or between 0 and 1 in fuzzy set. A benefit of fuzzy set theory in fuzzy decision tree is to improve the ability to understand when to use a decision tree quantitative attributes. By using fuzzy technique can improve the analysis when the classification of new cases [3]

Fuzzy decision tree is the formalization of the mapping expression of an attribute value in a classroom. Mapping attributes consist of interconnected nodes with two or more sub tree and leafs or a decision. The decision tree is a popular and practical approach in machine learning. This method is used to estimate the discrete values of the target function. Learning function represented by a decision tree. The main benefits of the decision tree approach are the ability to visualize a decision tree solution. This is done by following any path traversed in the decision tree. The discovery of relationships in the decision tree is expressed in a set of rules. This rule set can then be used to develop expert systems. Decision tree model is using recursive way divide-and-conquer to divide the data set into several parts so that the data that has the same class will be on the label. Decision tree classifying data by sorting the data in descending order from the root to the node tree leafs. Based decision tree induction algorithms that have been developed by the ID3 algorithm, developed by Quinlan. ID3 algorithm works in a symbolic domain. ID3 developed in various fields of research. ID3 is designed to accommodate data in a symbolic domain and the data finally reached on a single node leafs. ID3 algorithm is applied recursively for each child node until all the nodes have samples of each class. Fuzzy decision tree allows data to be traced simultaneously through a miraculous branch node with degree range [0 1].

Combine elements of fuzzy decision trees with symbolic and sub-symbolic approach. Fuzzy sets and fuzzy logic allows a modeling language that relate to the uncertainty over the supply of a symbolic framework for knowledge is comprehensive. There are 3 difference fuzzy decision tree with crisp decision tree is:

1. Fuzzy decision tree is using fuzzy constraints for the separation criteria.
2. Fuzzy decision tree inference procedures have differences
3. Fuzzy set in fuzzy decision tree represent data that has been define.

Induction of fuzzy decision tree has two main components:

1. Procedure to build a fuzzy decision tree
2. Inference procedures for decision-making

The use of fuzzy decision tree allows the prediction of an object will have more than one class.

Information gain is a statistical value used to select attributes that will expand the tree and generate a new node in a fuzzy decision tree by ID3 algorithm. Entropy value is to define the value of information gain. Entropy value reflects the quality attributes as the branch attribute. Calculation of entropy and information gain can be describe as a collection of data sets with discrete attributes value conditions and a set of discrete valued decision attribute data. The set is a knowledge representation system

2. RESEARCH METHOD

In the pattern recognition necessary, clear determination of classification rules to be able to recognize a pattern well. Classification of data is need to identify potentially characteristics in the set N objects contain in a database and to categorize them into different groups. Stages in this study is begins with pre-processing (data acquisition), Process consist of enhancement and segmentation, Feature Extraction, calculate Information Gain, preparing Fuzzy Decision Tree and the last is the extraction of Fuzzy Rules for Classification. The method of producing batik image classification of rules is show in Figure 1.

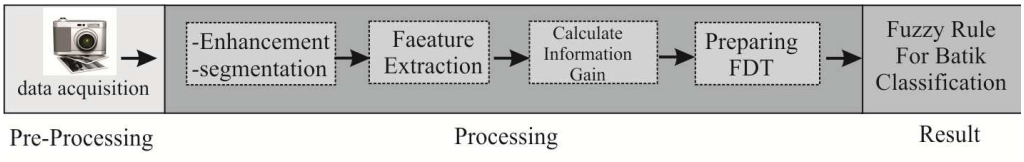

Figure 1. Method of producing batik image classification of rules

Texture is a certain regularity of the patterns formed by the arrangement of pixels in the image. Terms of texture formation is:

1. The existence of primitive patterns consist many pixels. These patterns are the basic elements of a form.

2. Previous pattern of primitive recur at intervals and directions so that the recurrence of certain predictable characteristics. Some features textures that will use in this study are entropy, energy, contras and homogeneity.

3. RESULTS AND ANALYSIS

This section will explain the process of extracting classification rules on batik pattern recognition. Amount of image data collected to determine the kinds of fabrics classified as batik or not batik. The amount of data consists of 10 types of fabric, as in Table 1.

Table 1. The data set of fabrics

Types Of Fabric	Entropy	Energy	Contras	Homogeneity	Recognition
	6.7865	223.8874	Low	yes	batik
	6.6987	205.4322	High	no	not batik
	6.3448	232.9982	low	Yes	not batik
	7.5195	237.9991	high	yes	batik
	5.342	208.9821	low	yes	not batik
	6.776	243.2123	low	no	batik
	6.4532	215.9879	high	Yes	not batik
	5.4934	223.9932	low	Yes	not batik
	7.6654	232.0021	Low	yes	batik
	7.1162	238.9899	low	no	batik

Each fabric has 4 feature values are entropy, energy, contrast and homogeneity. In the case, triangles are using to represent the membership function of fuzzy sets that computation can be efficiently computing. Experienced experts usually predefine membership functions. They also can be deriving through automatic adjustments [3]. Entropy and energy have three fuzzy regions low, middle and high. Thus, three fuzzy membership values have been produce for each course according to the predefined membership functions.

3.1. Calculate Information Gain

To represent a continuous fuzzy set, we need to express it as a function and then map the elements of the set to their degree of membership. Transform the quantitative values of each value into fuzzy sets. The transformation procedure is repeated for the others value. The result is show in Table 2.

Table 2. The data set of Batik Image in Fuzzy form

Types Of Fabric	Entropy	Energy	Contras	Homogeneity	Recognition
1	middle	middle	Low	yes	Batik
2	middle	low	High	no	no batik
3	middle	middle	low	Yes	no batik
4	high	high	high	yes	Batik
5	Low	low	low	yes	no batik
6	middle	high	low	no	Batik
7	middle	low	high	Yes	no batik
8	Low	middle	low	Yes	no batik
9	high	middle	Low	yes	Batik
10	high	high	low	no	Batik

Form knowledge representation system

$$J = \{U, C \cup D\}, U = \{1, \ldots, 10\}, C = \{Ent, Cont, Ener, Hom\}, D = \{Recognition\}$$

The class label is recognition, has two distinct values {batik, not batik}. There are two distinct classes (m=2), let class d_1 represents 'batik' and class d_2 represents 'not batik'. There are 7 samples of class

'batik' and 3 samples of class 'not batik'. We can calculate the information gain with the formula $I(S_1, \ldots, S_m) = -\sum_{i=1}^{m} P_i \, Log_2 P_i$. Compute the entropy for each feature value, for contrast, it has three distinct value {high, middle, low}, U can be partitioned into three subsets {s_1, s_2, s_3}. The result of computing the value of entropy can compute the information gain with the formula $Gain(c_i) = I(S_{1j}, \ldots, S_{mj}) - E(c_i)$. From the computation we get the information gain for Gain (Ent) = 0.4, Gain (Cont) = 0.6, Gain (Ener) = 0.33 and Gain (Hom) = 0.05. Gain contrast has highest information gain among the others, so, Contrast is select as the feature value to split the tree.

3.2 Constructing a Decision Tree

Contrast is selected feature to construct teh decision tree. The equivalences classes: high {4, 6, 10}, middle {1, 3, 8, 9} and low {2, 5, 7}. The subset class middle needs to further split. Then the whole decision tree has been complete as figure 2.

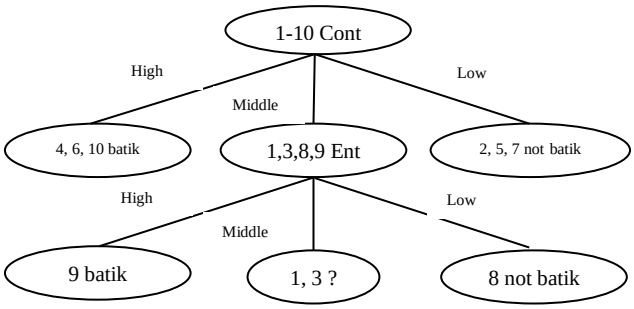

Figure 2. Decision Trees based on Information Gain

3.3 Extract Batik Image classification Rule

Classification if-then rules extracted from the equivalence classes. For equivalences class {4, 6, 10}, those samples all have the identical feature values: Cont=high, Recognition=batik. Therefore, the condition feature values (cont=high) as the rule antecedent and use the class label attribute value (Recognition = batik) as the rule consequent, we can get the following classification rule:

IF Cont="high" THEN Recognition="batik" Similarly, the other classification rules can be extract at this manner. The rules as follows:

1. IF cont='high' THEN recognition='batik'
2. IF cont='low' THEN recognition='not batik'
3. IF cont='middle' AND Ent='high' THEN recognition='batik'
4. IF cont='middle' AND Ent='low' THEN recognition='not batik'

3. CONCLUSION

This paper is concerned with fuzzy sets and decision tree. It present a decision tree model based on fuzzy set theory. Fuzzy set theory applied to transform real world data into fuzzy linguistic form and information theory to construct a decision tree. Finding the best split point and performing the split are the main task in decision tree induction method. Through the fuzzy decision tree, it can make Extract classification rules that can classify patterns of batik and batik pattern instead and make classification task difficult to become possible.

REFERENCES

[1] S. Tsang, B. Kao, KY Yip, WS. Ho and SD Lee, "Decision Trees for Uncertain Data. *IEEE Transactions on Knowledge and Data Engineering* [Internet]. 2011 Jan;23(1):64–78.

[2] Z. Juanjuan, LU Huijun, LI Yue and C. Junjie, "A kind of Fuzzy Decision Tree Based on the Image Emotion Classification". 2012 *International Conference on Computing, Measurement, Control and Sensor Network*. 2012;(60970059):167–70.

[3] T. Wang and H. Lee, "Constructing a Fuzzy Decision Tree by Integrating Fuzzy Sets and Entropy". *Proceeding ACOS'06 Proceedings of the 5th WSEAS international conference on Applied computer science*. 2006;306–11.

[4] FX. Aymerich, J. Alonso, ME Cabañas, and M. Comabella, Decision tree based fuzzy classifier of 1 H magnetic resonance spectra from cerebrospinal fluid samples. *Fuzzy Sets and Systems* [Internet]. Elsevier; 2011;170(1):43–63.

[5] J. Hu, S. An, J. Wang, Fuzzy Decision Tree Algorithm for Customer Knowledge Management of PLM. 2011 *International Conference of Information Technology, Computer Engineering and Management Sciences* [Internet]. IEEE; 2011 Sep [cited 2013 Apr 29];(2):139–42.

BIBLIOGRAPHY OF AUTHORS

	Received the master's degree in computer science from Gadjah Mada University in 2010. She is a lecture at the Department of Electronic Engineering, Tadulako University, Palu since 2005. Her research interest in fuzzy logic, artificial intelligent and pattern recognition. Now, she is a doctoral student in Department of Computer Science and Electronics Instrumentation, Faculty of Mathematic and Natural Science, Gadjah Mada University
	Received the doctor's degree in computation from Computation, University of Manchester Institute of Science and Technology, UK in 1996 and the master's degree from Computer Science, the University of Manchester, UK in 1990. He is a lecturer at the Detartment of Mathematics, Gadjah Mada University, since 1982 and a lecturer at the Department of Computer Science, Gadjah Mada University, since 1990. His research interest in Database Management Systems, Operating Systems, Management Information Systems, Fuzzy Logics and Software Engineering.