

## CLASSIFIER LEARNING ALGORITHM BASED ON GENETIC ALGORITHMS

LIYAN DONG<sup>1</sup>, HAILONG ZHANG<sup>1</sup>, XIANGSHI REN<sup>2</sup> AND YONG-LI LI<sup>3</sup>

<sup>1</sup>College of Computer Science and Technology  
Jilin University  
Changchun 130012, P. R. China  
dongliyan@gmail.com

<sup>2</sup>The School of Information  
Kochi University of Technology  
Kochi 782-8502, Japan  
ren.xiangshi@kochi-tech.ac.jp

<sup>3</sup>College of Computer  
Northeast Normal University  
Changchun 130021, P. R. China

Received April 2008; revised June 2009

**ABSTRACT.** *Combining the merits of the optimal search in Bayesian Theory and GA, this paper improves the TAN classifier, which maintains the classification accuracy rate of TAN and gives a restricted BAN classifier learning algorithm, namely GBAN. Learning the structure of the BAN classifier, a fitness function based on logarithm likelihood and the corresponding genetic operator are designed, and network structure code scheme is also designed, so that GBAN can constrict to global optimal structure. With the use of several data sets from the UCI Database, the comparisons of the classification results between GBAN and other algorithms such as TAN, BAN, GBN and Naive Bayes are given, respectively. And the results show that GBAN algorithm can get a better classification performance.*

**Keywords:** Bayesian network classifier, Genetic algorithm, Algorithm likelihood, Classification

1. **Introduction.** Naive Bayesian Classifier is one of the simple and efficient probabilistic ways of classifying, which shows very good effects under the assumption of “class conditional independence” [1,4], that is, given the class label of a tuple, the values of the attributes are assumed to be conditionally independent of one another. Unlike other ways of classifying, this assumption simplifies computation. To get parameter of each attribute, only the values’ frequencies of each attribute in the training set need to be calculated and there is no need for searching. But, in the real world, it’s hard to meet the assumption of conditional independence; so many people are trying to find some other models that can work efficiently without this assumption. TAN classifier (Tree Augmented Naive Bayes) [13] can obtain MWST (maximum weight span tree) based on Chow-Liu’ MWST algorithm, which loosens the assumption of class conditional independence and keeps the structural simplicity and advantage. But the accuracy of TAN classifier decreases when relationships among the attributes of dataset become more complex. BAN classifier (Bayes Network Augmented Naive-Bayes) [4] loosens the assumption further, and shows stronger learning capacity when the relationships among the attributes are complex. However, the process of BAN learning is essentially a constrained learning process of Bayesian Networks which has several disadvantages such as searching in very