Available online at www.sciencedirect.com
SCIENCE DIRECT®
Information Sciences 169 (2005) 205–226
INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL
www.elsevier.com/locate/ins

Application of fuzzy Naive Bayes and a real-valued genetic algorithm in identification of fuzzy model ☆

Yongchuan Tang [a,b,*], Yang Xu [b]

[a] *College of Computer Science, Zhejiang University, Hangzhou, Zhejiang 310027, PR China*
[b] *Department of Applied Mathematics, Southwest Jiaotong University, Chengdu, Sichuan 610031, PR China*

Received 5 January 2003; received in revised form 8 March 2004; accepted 12 May 2004

Abstract

We present a method to identify a fuzzy model from data by using the fuzzy Naive Bayes and a real-valued genetic algorithm. The identification of a fuzzy model is comprised of the extraction of "if–then" rules that is followed by the estimation of their parameters. The involved parameters include those which determine the membership function of fuzzy sets and the certainty factors of fuzzy if–then rules. In our method, as long as the fuzzy partition in the input–output space is given, the certainty factor of each rule is computed with the fuzzy conditional probability of the consequent conditioned on the antecedent by using the fuzzy Naive Bayes, which is a generalization of Naive Bayes. The fuzzy model involves the rules characterized by the highest values of certainty factors. The certainty factor of each rule is the fuzzy conditional probability, and it reflects the inner relationship between the antecedent and the consequent. In order to improve the accuracy of the fuzzy model, the real-valued genetic algorithm

☆ This work has been supported by the National Natural Science Foundation of China (Grant No. 60074014) and Chinese 973 project (Grant No. 2002CB312106).

* Corresponding author. Address: College of Computer Science, Zhejiang University, Hangzhou, Zhejiang 310027, PR China. Tel./fax: +86 0571 87952639.

E-mail address: yongchuan@263.net (Y. Tang).

doi:10.1016/j.ins.2004.05.004

is incorporated into our identification process. This process concerns the optimization of the membership functions occurring in the rules. We just involve the parameters of membership function of the fuzzy sets into the real-valued genetic algorithm, since the certainty factor of each rule can be computed automatically. The performance of the model is shown for the backing-truck problem and the prediction of Mackey–Glass time series.

Keywords: Fuzzy model; Fuzzy Naive Bayes; Real-valued genetic algorithm; Conditional probability; Backing-truck; Time series prediction

1. Introduction

When some additional parameters are incorporated into fuzzy modelling, the power of knowledge representation will be strengthened greatly, since in the application, the knowledge acquisition often involves imprecision [25]. The knowledge imprecision may be derived from several sources: the knowledge acquisition process used, the availability of the domain experts, the knowledge representation and reasoning method being employed. Therefore, it is necessary to introduce some additional parameters. A popular parameter in fuzzy modelling is the certainty factor of each fuzzy if–then rule, which describes how certain the relationship between the antecedent and the consequent of this rule is. The certainty factors can also fine tune the final fuzzy model, since the influence of the certainty factor is local [16]. Although the advantage of the certainty factor is clear, its determination is still a difficult problem. In general, the identification of the fuzzy model is a key problem in the application. There are mainly two approaches to tackle this problem. One is directly summarizing the operators' or experts' experiences and translating their knowledge into fuzzy rules. The knowledge acquisition and verification processes, however, are difficult and time-consuming. The other approach is obtaining fuzzy rules through machine learning, with which knowledge can be automatically generated or extracted from sample cases or examples.

In previous research about the automatic extraction of fuzzy models from data, the certainty factors were independently determined [1,2]. This means that these works didn't take into account the explicit relationship between the certainty factors and the fuzzy partition in the input–output space. The omission of this relationship will cause the number of parameters which should be determined independently to increase, and fail to give the theoretical instruction in designing the fuzzy model. One possible way to discuss the relationship between the certainty factors and other parts of the fuzzy model is the fuzzy Naive Bayes presented in Section 3.3.