ON THE INDUCTIVE MODEL GENERATION

Strijov, V.

Computing Centre of the Russian Academy of Sciences, Moscow, Russia

vadim@strijov.com

This talk is devoted to the problem of the automatic model creation in regression analysis. The models are intended for dynamic systems behavior analysis in biology and medicine. The theory and the practice of the inductively generated models will be examined.

The problem motivation is the following. Mathematical modelling in biology has two issues: first, to create a model of a dynamic biosystem using the expert knowledge and second, to discover a model and knowledge using the measured data. So there are the model-driven and the data-driven approaches, and each one has its own strengths and weaknesses. The first one gives models that could be interpreted by experts in a field of application but usually they have poor prediction quality. The second one gives models of good quality but often too complex and non-interpretable by experts. The suggested approach gathers strong sides of these two: the result the model could be explained and it relies on the measured data. It allows getting the model with fair quality and generalization ability in comparison to universal models [1].

A model is selected from an inductive generated set of the trial models according to the notion of adequacy: the model must be simple, stable and precise. These criterions are target functions and they are assigned according to given data. It is supposed that given data carries the information on the searched model and the noise as well. The hypothesis of the probability distribution function defines a data generation hypothesis and as follows, the target functions.

The outline of the automatic model creation is the following. A sample data, which consist of several independent variables and one dependent variable are given. Experts makes set of terminal functions. These models are arbitrary superposition, inductively generated using terminal functions. Experts could also make initial models for inductive modification. When generated models are tuned, a model of the optimal structure is selected. The methods of inductive superposition generation and model selection will be discussed. This project is supported by RFBR, grant 07-07-00181.

References

1. Strijov, V. The search of a parametric regression model in an inductive-generated set / Journal of Computational Technologies. 2007. No 1. P. 93-102.