Gaussian Processes for Machine Learning

Dejan Petelin[1,2]

[1] Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia
[2] Jožef Stefan International Postgraduate School (New Media and E-science, 1st-year)

dejan.petelin@ijs.si

Gaussian process (GP) models form a new, emerging complementary method for nonlinear system identification. The GP model is a probabilistic nonparametric black-box model. It differs from most of the other frequently used black-box identification approaches as it does not try to approximate the modeled system by fitting the parameters of the selected basis functions, but rather searches for the relationship among measured data. Gaussian processes models are closely related to approaches such as Support Vector Machines, and specially Relevance Vector Machines [1]. Because GP model is a Bayesian model, the output of Gaussian process model is a normal distribution, expressed in terms of mean and variance. Mean value represents the most likely output and the variance can be viewed as the measure of its confidence. Obtained variance, which depends on amount of available training data, is important information distinguishing the GP models from other non-bayesian methods. Gaussian process can be used for model identification when data are heavily corrupted with noise, and when there are outliers or gaps in the training data. Another useful attribute of the GP model is the possibility to include various kinds of prior knowledge into the model, e.g. local models, static characteristic, etc.

A noticeable drawback of the system identification with Gaussian process models is computation time necessary for modeling. Gaussian process regression involves several matrix computations which load increases with the third power of the number of training data, such as matrix inversion and the calculation of the log-determinant of used covariance matrix. This computational greed restricts the number of training data, to at most a few thousand cases.

This limitation precludes usage of GP models for many real applications, usually dynamic systems, where the amount of data constantly rises. In these cases, the ability of online learning of GP models with constant time of an update is needed. This could be achieved with maintenance of a set of the most informative data. The size of a set should be small enough that an update could be completed before new data arrives. In many cases new data is much more important than old data. In these cases the feature of forgetting old data is used.

References:

[1] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006
[2] J. Kocijan, D. Petelin, V. Tanko, Identifikacija modelov črne skrinjice na podlagi Gaussovih procesov, IJS delovno poročilo, 2010
[3] L. Csató and M. Opper. Sparse On-Line Gaussian Processes. Morgan Kaufmann, *Neural Computation*, 14(3):641-669, 2002.
[4] D. Petelin. Sprotno učenje modelov na podlagi Gaussovih procesov. Diplomsko delo, 2009.