Gaussian Process Models for Systems Identification

Juš Kocijan

Abstract—Different models can be used for nonlinear dynamic systems identification and the Gaussian process model is a relatively new option with several interesting features: model predictions contain the measure of confidence, the model has a small number of training parameters and facilitates structure determination, and different possibilities of including prior knowledge exist. In this paper the framework for the identification of a dynamic system model based on the Gaussian processes is presented and a short survey with a comprehensive bibliography of published works on application of Gaussian processes for modelling of dynamic systems is given.

I. INTRODUCTION

While there are numerous methods for the identification of linear dynamic systems from measured data, the nonlinear systems identification requires more sophisticated approaches. The most common choices include artificial neural networks, fuzzy models and others. Gaussian process (GP) models present a new, emerging, complementary method for nonlinear system identification.

The GP model is a probabilistic, non-parametric black-box model. It differs from most of the other black-box identification approaches as it does not try to approximate the modelled system by fitting the parameters of the selected basis functions but rather searches for the relationship among measured data. Gaussian process models are closely related to approaches such as Support Vector Machines and specially Relevance Vector Machines [3].

The output of the Gaussian process model is a normal distribution, expressed in terms of mean and variance. The mean value represents the most likely output and the variance can be interpreted as the measure of its confidence. The obtained variance, which depends on the amount and quality of available identification data, is important information distinguishing the GP models from other methods. The GP model structure determination is facilitated as only the covariance function and the regressors of the model need to be selected. Another potentially useful attribute of the GP model is the possibility to include various kinds of prior knowledge into the model, see [79].

The GP model was first used for solving a regression problem in the late seventies, but it gained popularity within the machine learning community in the late nineties of the twentieth century. Results of a possible implementation of the GP model for the identification of dynamic systems were presented only recently, e.g. [11], [54]. The investigation of the model with uncertain inputs, which enables the propagation of uncertainty through the model, is given in [20], [33], [39] and illustrated in [27], [47] and many others.

The purpose of this paper is twofold. First, to present the procedure of dynamic system identification using the model based on Gaussian processes taken from [83]. Second, a comprehensive bibliography of published works on Gaussian processes application for modelling of dynamic systems with a short survey is given.

Many of dynamic systems are often considered as complex, however simplified input/output behaviour representations are sufficient for certain purposes, e.g. feedback control design, prediction models for supervisory control, etc. In the paper it is explained how the advantages of Gaussian process models can be used in identification and validation of such models.

The paper is organised as follows. In Section 2 basic principles of the GP model and its use in dynamic system identification are described. The methodology of the identification with a GP model is given in Section 3. Section 4 contains a short survey of comprehensive bibliography on GP modelling of dynamic systems. In the last section the discussion and main conclusions are gathered.

II. MODELLING OF DYNAMIC SYSTEMS WITH GAUSSIAN PROCESSES

A. Modelling with the GP model

Here, modelling with the GP model is presented only in brief, for a more detailed explanation see e.g. [79].

A Gaussian process is a Gaussian random function, fully described by its mean and variance. Gaussian processes can be viewed as a collection of random variables \( f(x_i) \) with joint multivariate Gaussian distribution: \( f(x_1), \ldots, f(x_n) \sim \mathcal{N}(0, K) \). Elements \( K_{ij} \) of the covariance matrix \( K \) are covariances between values of the function \( f(x_i) \) and \( f(x_j) \) and are functions of corresponding arguments \( x_i \) and \( x_j \): \( K_{ij} = C(x_i, x_j) \). Any function \( C(x_i, x_j) \) can be a covariance function, providing it generates a nonnegative definite covariance matrix \( K \).

Certain assumptions about the process are made implicitly with the covariance function selection. The stationarity of the process results in the value of covariance function \( C(x_i, x_j) \) between inputs \( x_i \) and \( x_j \) depending only on their distance and being invariant to their translation in the input space, see e.g. [79]. Smoothness of the output reflects in outputs...
The Gaussian covariance function: $C(x_i, x_j) = \text{cov}[f(x_i), f(x_j)]$

\[= v \exp \left[ -\frac{1}{2} \sum_{d=1}^{D} w_d (x_i^d - x_j^d)^2 \right] + \delta_{ij} \nu_0 \]

where $D$ is the length of the vector $x$ and $\Theta = [w_1, \ldots, w_D, \nu_0]^T$ is a vector of parameters called hyperparameters.\(^1\) The first term in (1) corresponds to functional dependence under presumed stationarity, while the second term corresponds to noise. Hyperparameter $v$ controls the magnitude of the covariance of regressors $w_i$ representing the relative importance of each component $x^d$ of vector $x$. The part $\delta_{ij} \nu_0$ represents the covariance between outputs due to white noise, where $\delta_{ij}$ is the Kronecker operator and $\nu_0$ is the white noise variance. When assuming different kinds of noise the covariance function should be changed appropriately, e.g. [8]. With the use of covariance function (1) the total number of the GP model parameters is $D + 2$ for the size $D$ input, where for example the number of comparable artificial neural networks parameters would be considerably larger.

The GP model fits nicely into the Bayesian modelling framework. The idea behind GP modelling is to place the artificial neural networks parameters would be considerably for the total number of the GP model parameters is $D + 2$ for the size $D$ input, where for example the number of comparable artificial neural networks parameters would be considerably larger.

The presented GP model was originally used for modelling static nonlinearities, but it can be extended to model dynamic systems as well [39], [54], [3]. Our task is to model the dynamic system (2), where $x = [y(k-1), \ldots, y(k-L), u(k-1), \ldots, u(k-L)]$ (10) is the vector of regressors that determines nonlinear ARX model structure and be able to make multi-step ahead model prediction.

\[^1\]The parameters of a Gaussian process are called hyperparameters due to their close relationship to the hyperparameters of a neural network [79].
A demonstration of a Gaussian process model response is given in Figure 2.

III. GAUSSIAN PROCESS MODEL IDENTIFICATION METHODOLOGY

In this section the framework for dynamic system identification with GP models taken from [83] is given. The identification framework consists of roughly six stages:

- defining the purpose of the model,
- model selection,
- design of the experiment,
- realisation of the experiment and data processing,
- training of the model and
- model validation.

The model identification is an iterative process. Returning to some previous procedure step is possible at any step in the identification process and is usually necessary.

A. The model purpose and model selection

The decision for the use of a specific model derives from the model purpose and from the limitations met at the identification process. In this paper selection of the GP model is presumed. This approach can be beneficial when the information about the system exists in the form of input/output data, when data are corrupted, e.g. by noise and measurement errors, when a measure of confidence in model prediction is required and when there is a relatively small amount of data in respect to the selected number of regressors.

After the model is selected, its structure must be determined next. In the case of the GP model this means selecting the covariance function and the model regressors. The choice of the covariance function reflects the relationship between data and is based on prior knowledge of the process. The standard choice for smooth and stationary processes is function (1). Prior knowledge about other attributes, e.g. periodicity, non-stationarity, can be expressed through a different choice of the covariance function [79].

The second part of structure determination is the choice of proper regressors. In the case of a dynamic system model this also means selecting the model order, which is the area of intensive research, as it is common to all nonlinear identification methods.

The most frequent approach for regressor selection is the so called validation based regressor selection, where the
search for the optimal vector of regressors is initiated from some basic set of regressors. After the model optimisation and cross-validation, the regressors are added to or taken from the model. Prospering models according to selected performance are kept while dissatisfying models are rejected. In the case of normalised inputs the influence of each regressor can be observed through the value of the associated hyperparameter. If the associated regressor is not relevant enough it can be removed from the perspective model.

B. Obtaining data – design of the experiment, experiment and data processing

Data describing the unknown system is very important in any black-box identification. For a good description of the process the influential variables and proper sample time must be chosen.

The design of the experiment and the experiment itself are, as is always the case in systems modelling, very important parts of the identification procedure. The quality of the model depends on the system information contained in the measurement data, regardless of the identification method. Nevertheless, the design of the experiment is not the focus of this paper.

As already mentioned the Gaussian process modelling approach relies on the relation among input/output data and not on approximation with basis functions. Consequently, this means that the distribution of identification data within the process operating region is crucial for the quality of the model. Model predictions can be informative only if the inputs to the model lie in the regions, where training data is available. The GP model is good for interpolation, but not for extrapolation, which is indicated by large variances of model predictions.

Consequently, the data for model training should be chosen reasonably, which can be obstructed by the nature of the process (e.g. limitations in the experiment design in industrial processes, physical limitations of the system). The preprocessing of measured data, such as normalisation to cancel the influence of different measuring scales, can be pursued.

C. Model training

In the GP model approach training means optimization of hyperparameters \( \Theta \) from (1). Each hyperparameter \( w_d \) expresses the relative importance of the associated regressor, similar to the automatic relevant detection (ARD) method [79], where a higher value of \( w_d \) expresses higher importance of the regressor. Hyperparameter \( v \) expresses the overall scale of correlations and hyperparameter \( v_{0} \) accounts for the influence of noise. Several possibilities of hyperparameter determination exist. A very rare possibility is that hyperparameters are known in advance as prior knowledge. Almost always, however, they must be determined from the training data, where different approaches are possible, e.g. [39]. Mostly the likelihood maximization (ML) approach is used as it gives good results despite its simplification, where any optimization method could be used to achieve ML [39].

D. Model validation

Validation concerns the level of agreement between the mathematical model and the system under investigation [2] and it is many times underemphasised despite its importance. Several features can represent the quality of the model. Their overview can be found e.g. in [2], [1]. The most important are model plausibility, model falseness and model purposiveness, explained as follows.

Model plausibility expresses the model’s conformity with the prior process knowledge by answering two questions: whether the model “looks logical” and whether the model “behaves logical”. The first question addresses the model structure, which in the case of GP models means mainly the plausibility of the hyperparameters. The second one is concerned with the responses of the model output to typical events on the input, which can be validated with visual inspection of the responses as is the case with other black-box models.

Model falseness reflects the agreement between the process and the model output or the process input and the output of the inverse model. The comparison can be done in two ways, both applicable to GP models: qualitatively, i.e. by visual inspection of differences in responses between the model and the process, or quantitatively, i.e. through evaluation of performance measures. Beside commonly used performance measures such as e.g. mean squared error \( \text{MSE} \) and mean relative square error (MRSE, which compares only the mean prediction of the model to the output of the process):

\[
\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} e_i^2
\]  

\[
\text{MRSE} = \sqrt{\frac{\sum_{i=1}^{N} e_i^2}{\sum_{i=1}^{N} y_i^2}}
\]  

where \( y_i \) and \( e_i = \hat{y}_i - y_i \) are the system’s output and prediction error in \( i \)-th step of simulation, the performance measures such as log predictive density error (LD, [39], [54]) can be used for evaluating GP models, taking into account not only mean prediction but the entire predicted distribution:

\[
\text{LD} = \frac{1}{2} \log(2\pi) + \frac{1}{2N} \sum_{i=1}^{N} \left( \log(\sigma_i^2) + \frac{e_i^2}{\sigma_i^2} \right)
\]  

where \( \sigma_i^2 \) is the prediction variance in \( i \)-th step of simulation.

Performance measure LD weights the prediction error \( e_i \) more heavily when it is accompanied with smaller predicted variance \( \sigma_i^2 \), thus penalising overconfident predictions more than acknowledged bad predictions, indicated by higher variance. Another possible performance measure, applicable in the training procedure, is the negative log-likelihood of the training data (LL, [39]):

\[
\text{LL} = -\frac{1}{2} \log | \mathbf{K} | + \frac{1}{2} \mathbf{y}^T \mathbf{K}^{-1} \mathbf{y} + \frac{N}{2} \log(2\pi),
\]  

where \( \mathbf{K} \) is the covariance matrix, \( \mathbf{y} \) is the vector of targets and \( N \) is the number of training points. LL is the measure
inherent to the hyperparameter optimisation process, see (5), and gives the likelihood that the training data is generated by given, i.e. trained, model. The smaller the MRSE, LD and LL are, the better the model is.

Variance of the model predictions on a validation signal can be a validation measure itself, as it indicates whether the model performs well in regions where identification data were available. Nevertheless, it should be used carefully and in combinations with other validation tools, as predictions with small variance are not necessary good.

Model purposiveness or usefulness tells whether or not the model satisfies its purpose, which means the model is validated when the problem that motivated the modelling exercise can be solved using the obtained model. Here, again, the prediction variance can be used, e.g. when the prediction confidence is too low, the model can be labelled as not purposive.

IV. SURVEY OF PUBLICATIONS ON GAUSSIAN PROCESS MODELS OF DYNAMIC SYSTEMS

The GP model was first used for solving a regression problem in the late 1970s, but it only gained popularity within the machine-learning community in the late 1990s. Furthermore, the results of a possible implementation of the GP model for the identification of dynamic systems were presented as recent as the last decade.

After what can be described as initial publications in year 1999 [4], year 2000 [5], [6] and year 2001 [7], [8], numbers of publications start to grow. Numerous publications on conferences and as internal, but publicly available publications occurred in years 2002 [9]-[19], 2003 [20]-[37] and 2004 [38]-[45]. After the first journal publication in year 2003 [24], publications in years 2005 [46]-[68], 2006 [69]-[81] and 2007 [82]-[97] contain more versatile publications including journal papers, book chapters and books mentioning use of GP models for the modelling of dynamic systems. In spite of efforts to be very thorough it is possible that the list of publications until year 2007 is not complete, but it certainly represents the majority of publications on Gaussian process models of dynamic systems.

These publications have explored use of Gaussian process models for various applications:

- dynamic systems modelling, e.g., [10],[11],[27],[65]
- time-series prediction, e.g., [7],[73],
- fault detection, e.g., [12],[13],[18],[55],
- smoothing, e.g., [82],
- etc.

The utility to provide the information about the model prediction confidence made Gaussian process models attractive for modelling case studies in various domains like: chemical engineering [91] and process control [93], biomedical engineering [84], biological systems [83], environmental systems [73], power systems [43] and engineering [60], motion recognition [65], etc., to list just a few. It is worth noticing that the utility of Gaussian process modelling could be interesting also for use in other domains and applications therein.

V. CONCLUSIONS

In this paper it is explained how the Gaussian process model is used for dynamic systems identification with emphasis on some of its properties: model predictions containing the measure of confidence, low number of parameters and facilitated structure determination.

The prediction variance is one of the main differences between the GP model and other black box models. It can be effectively used in the usefulness validation, where the lack of confidence in the model prediction can serve as the grounds to reject the model as not useful. The prediction variance can also be used in falseness validation, whether via specific performance measures such as log-predictive density error, or through observation of confidence limits around the predicted output. Despite its usefulness in model validation, it should be accompanied with standard validation tools, as the small variance does not necessarily mean that the model is of good quality.

In the validation based regressor selection procedure the log-predictive density error and the log-likelihood of the training data can be useful in selecting model regressors. In the case of normalised inputs, the model hyperparameters indicate the influence of corresponding regressors and can be used as a tool for removal of non-influential regressors at the regressor selection stage of the model selection.

Small amounts of data relative to the number of selected regressors, data corrupted with noise and measurement errors and the need for the measure of model prediction confidence could be the reasons to select identification with the GP model. If there is not enough data or it is heavily corrupted with noise, even the GP model cannot perform well, but in that case the inadequacy of the model and the identification data is indicated through higher variance of the predictions.

The short survey and bibliography on Gaussian process models for dynamic systems shows that the interest in this modelling approach and its applications is growing. Published results have shown the GP model’s potential for the identification of nonlinear dynamic systems and where the advantages of the GP model could be effectively used, e.g., for control design, diagnostic system design etc.

VI. ACKNOWLEDGMENTS

The author gratefully acknowledges the support of the Slovenian Research Agency Grant No. P2-0001.

REFERENCES

Published references on Gaussian process models of dynamic systems and their publications listed by publishing year

1999


2000


2001


2002


2003


2004


2005


2006


2007


AUTHORS BIOGRAPHY

Juš Kocijan received the doctor degree in electrical engineering from the Faculty of Electrical Engineering, University of Ljubljana. He is currently a senior researcher at the Department of Systems and Control, Jozef Stefan Institute, Ljubljana and Professor of Electrical Engineering at the School of Engineering and Management, University of Nova Gorica, Slovenia. His main research interests are: applied nonlinear control and multiple model and probabilistic approaches to modelling and control. Prof. Kocijan serves on a Board of Editors of IFAC journal Engineering Applications of Artificial Intelligence and on Editorial Advisory Boards of Recent Patents on Electrical Engineering and The Open Automation and Control Systems Journal. He is a Senior member of IEEE, a member of Automatic control society of Slovenia and Slovenian Society for Simulation and Modelling.