Automatic Summarization of Composite Nonparametric Time-Series Models

James Robert Lloyd
University of Cambridge
jrl44@cam.ac.uk

David Duvenaud
University of Cambridge
dkd23@cam.ac.uk

Roger Grosse
M.I.T.
rgrosse@mit.edu

Joshua B. Tenenbaum
M.I.T.
jbt@mit.edu

Zoubin Ghahramani
University of Cambridge
zoubin@eng.cam.ac.uk

Abstract

To complement automatic model-building and search methods, we demonstrate an automatic model-summarization procedure. Given a composite model, our method automatically constructs a report which visualizes and explains in words the meaning and relevance of each component of the model. These reports enable human model-checking, understanding of complex modeling assumptions, and understanding of complex structure in data.

1 Introduction

Parametric models such as linear regression usually have simple interpretations, and well-established methods for model-checking. In constrast, complicated non-parametric models may be reasonably viewed with suspicion, since their assumptions may be difficult to check, and their predictive implications difficult to explain.

Advances in nonparametric and compositional modeling have allowed fully automatic model search and construction of complex, structured, nonprametric regression models [1]. Such automatically-constructed models can be used for extrapolation without further human intevention, but for model-checking or dataset exploration, extrapolations alone may be insufficient.

This paper describes a procedure for automatically generating a human-readable report, examining and explaining the properties of a given nonparametric model on a given dataset. The supplementary material to this paper is a pair of reports completely automatically generated by our method. These reports summarize and describe datasets modeled by a Gaussian process with a complicated kernel. The different components of the kernel correspond to different features of the dataset, and the report discusses in detail the properties and significance of each component.

2 Gaussian Process Structure Search

Gaussian process models[2] use a kernel to define the covariance between any two function values: $\text{Cov}(y, y') = k(x, x')$. The covariance function, or *kernel*, specifies which structures are likely under the GP prior, which in turn determines the generalization properties of the model.

Different kernels can express wide varieties of statistical structures, such as local smoothness, periodicity, and changepoints. Products of kernels can express even richer structure, such as functions which are periodic only locally, or functions whose amplitude grows over time. The Gaussian process structure search (GPSS) procedure simply searches over sums and products of a simple set of base kernels, in order to find complex structure in time-series or multidimensional functions [1]. In our experiments, we used 5 base kernels: squared-exponential (SE), Periodic (PER), Linear (LIN), changepoint, and white noise. When given a dataset, GPSS produces a compositional kernel which can be distributed into a sum of products of these base kernrels. One property of Gaussian process models is that a sum of kernels is equivalent to a sum of functions. This means that the models produced by GPSS can typically be decomposed into a sum of parts.

3 Automatic Report Generation

Our procedure takes as its starting point a dataset and a composite kernel, which together define a joint probability distribution over a sum of functions. The procedure summarizes properties of this complex distribution to the user through a comprehensive report. These reports are designed to be intelligible to non-experts, to illustrate the assumptions made by the model, to illustrate the model's posterior uncertaintly, and most importantly, to enable model-checking by humans.

Our procedure produces a report with three sections: an executive summary, a detailed discussion of each component, and a section discussing plausible extrapolations.

Automatic English Summarization Because the individual components produced by our procedure are produced by an open-ended grammar, it is impossible to write special code to handle every type of component that GPSS might produce. This is possible even in an open-ended grammar of kernels because of the compositional nature of kernel structures. Each type of kernel in an expression modifies the form of the posterior in a different way. Table 1 gives some examples. Because the posterior distribution also depends on the data, our summarizer includes heuristics to describe the probable shape of the components being modeled, such as "smooth" or "monotonically increasing".

Source Code All GP hyperparameter tuning was performed by automated calls to the GPML toolbox[1]; Python code to perform all experiments is available on github.[2]

4 Example: Summarizing 400 Years of Solar Activity

To give an example of the capabilities of our procedure, we show excerpts from the report automatically generated on annual solar irradiation data from 1610 to 2011. This time series has two obvious features: a roughly 11-year cycle or solar activity, and a period lasting from 1645 to 1715 with much smaller variance than the rest of the dataset. This flat region corresponds to the Maunder minimum, a period in which sunspots were extremely rare [3].

4.1 Executive Summary

The first section of each report describes which structures were discovered in the model, and the relevant importance of the different components in explaining the data.

Figure 1 shows the automatically-generated summary of the solar dataset. The model discovers 9 seperate components which explain the data, and reports that the first 4 components explain 90% of the variance in the data. This might seem incongruous with the observation that there are two main features of the data, but if we examine the first four components, we see that the first component is explaining the mean of the dataset, the second explains the Maunder minimum, the third describes the long-term trend, and the fourth describes the 11-year period.

4.1.1 Signal versus Noise

One design challenge we encountered was seperating the recovered structure into signal and noise. Originally, the model always included a term corresponding to i.i.d. additive Gaussian noise. However, in practice, the distinction between signal and noise is unclear for two reasons. First, a component which varies arbitrarily quickly in time can be indistinguishable from noise. Second, the

[1] Available at www.gaussianprocess.org/gpml/code/

[2] Available at github.com/jamesrobertlloyd/gpss-research

The structure search algorithm has identified nine additive components in the data. The first 4 additive components explain 92.3% of the variation in the data as shown by the coefficient of determination (R^2) values in table 1. The first 8 additive components explain 99.2% of the variation in the data. After the first 5 components the cross validated mean absolute error (MAE) does not decrease by more than 0.1%. This suggests that subsequent terms are modelling very short term trends, uncorrelated noise or are artefacts of the model or search procedure. Short summaries of the additive components are as follows:

- A constant.
- A constant. This function applies from 1644 until 1713.
- A smooth function. This function applies until 1644 and from 1719 onwards.
- An approximately periodic function with a period of 10.8 years. This function applies until 1644 and from 1719 onwards.
- A rapidly varying smooth function. This function applies until 1644 and from 1719 on-

Figure 1: An example of an automatically-generated summary of a dataset. The dataset is broken into a set of diverse types of structure, and each structure is explained in simple terms.

Kernel	Description
SE	A smoothly varying function
SE × PER	A smoothly varying periodic function
SE × PER × LIN	A smoothly varying periodic function with growing amplitude

Table 1: A demonstration of how it is possible to write an open-ended summarizer. Different kernel structures modify the overall statistical structure of each component in independent ways.

variance of the noise may change over time (called heteroskedasticity), and this sort of pattern may be considered part of the signal. Because of the blurry distinction between signal and noise, we include a table which summarizes the relative contribution of each component in terms of held-out predictive power.

4.2 Decomposition plots

The second section of each report contains a detailed discussion of each component in turn. Every component is plotted, and various properties of the statistical structure represent are described. Because components with small variance are often not meaningful when plotted on their own, we also include plots of each component combined with all components of higher marginal variance.

Automatic Plotting In the component plots, each component's posterior is described in two ways. First, the posterior mean and variance of each component is plotted on its own. Second, the posterior mean and variance of all components shown so far is plotted against the data. By contrasting each of these plot with plots of earler components, we can see qualitatively how the current component is contributing to an overall explanation of the data.

The GPSS grammar allows any expression to be multiplied by a changepoint kernel, which encodes a local change in covariance. For example, foom about 1645 to 1715, solar activity decreased, and very few susnpots were observed, a period called the Maunder Minimum [3]. Figure 2 shows that GPSS captured this structure with a pair of changepoint kernels.

Figure 3 shows the isolated periodic component of the signal. Here we see one of the main benefits of isolating individual components: we can now see, by eye, extra structure that was not captured by the model. Specifically, we can see that the amplitude of the periodic component varies over time.

4.3 Extrapolation plots

The third section of each report shows extrapolations into the future, as well as posterior samples from each individual component of the model. These samples help to characterize the uncertainty

2.2 Component 2 : A constant. This function applies from 1644 until 1713

This component is constant. This component applies from 1644 until 1713.

This component explains 35.3% of the residual variance; this increases the total variance explained from 0.0% to 35.3%. The addition of this component reduces the cross validated MAE by 29.42% from 0.33 to 0.23.

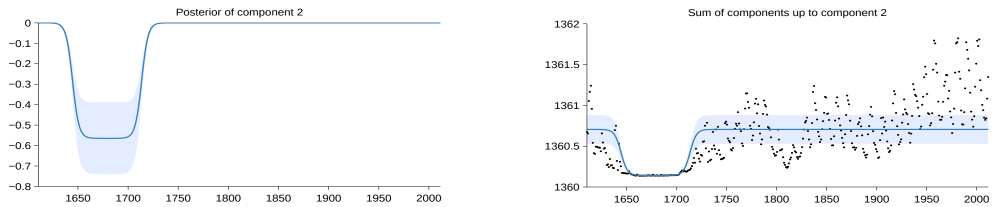

Figure 3: Posterior of component 2 (left) and the posterior of the cumulative sum of components with data (right)

Figure 2: Discovering the Maunder minimum. The kernel found by GPSS contained a pair of changepoints bracketing the period of low solar activity.

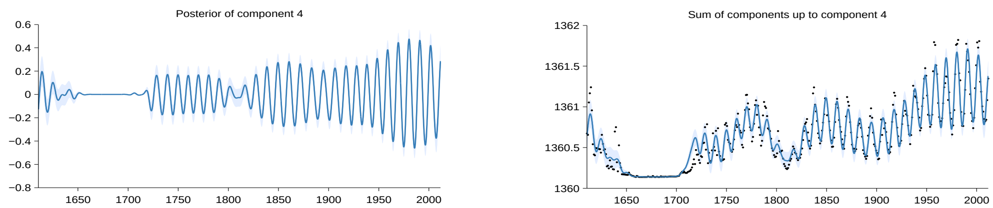

Figure 5: Posterior of component 4 (left) and the posterior of the cumulative sum of components with data (right)

Figure 3: Isolating the periodic component of the dataset. By isolating this aspect of the statistical structure, we can easily observe additional features, such as the shape of the peaks and troughs, or the fact that the amplitude changes over time.

captured by the model, and the extent to which different components contribute to predicting the future behavior of the time series. The predictive mean and variance of the signals shown in the summary plots are useful, but do not capture the joint correlation structure in the posterior. Showing posterior samples are a simple and universal way to illustrate joint statistical structure. For example, it is not clear from the left-hand plot in figure 4 whether or not the periodicity of the dataset is expected to continue into the future. However, from the samples on the right-hand size, we can see that this is indeed the case.

Extrapolating individual components We can also examine the model's expectations about the future behavior of individual components through sampling. Further plots in the extrapolation section show posterior samples for each individual additive component.

5 Discussion

In this paper, we demonstrated a method for automatically summarizing a compositional Gaussian process model. These summaries can enable human experts and non-experts to understand the implications of a model, check its plausibility, and notice structure not yet discovered by the model.

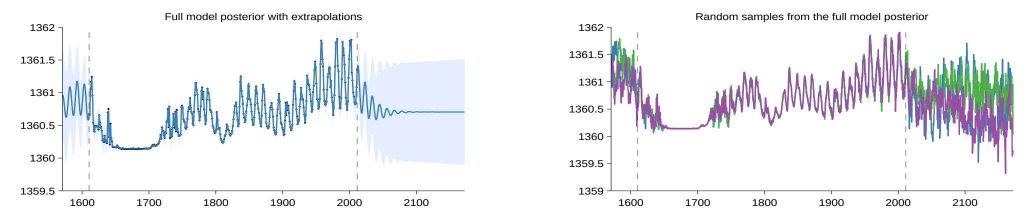

Figure 11: Full model posterior. Mean and pointwise variance (left) and three random samples (right)

Figure 4: Sampling from the posterior. These samples help show not just the predictive mean and variance, but also the predictive covariance structure. Note, for example, that the predictive mean (left) doesn't exhibit periodicity, but the samples (right) do.

References

[1] David Duvenaud, James Robert Lloyd, Roger Grosse, Joshua B. Tenenbaum, and Zoubin Ghahramani. Structure discovery in nonparametric regression through compositional kernel search. In *Proceedings of the 30th International Conference on Machine Learning*, June 2013.

[2] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, MA, USA, 2006.

[3] J. Lean, J. Beer, and R. Bradley. Reconstruction of solar irradiance since 1610: Implications for climate change. *Geophysical Research Letters*, 22(23):3195–3198, 1995.