
Predictive Model for Time-Series Data with Bayesian Non-parametrics

Chawannut Prommin, cp626

Serena Li, sl2327

Yutao Han, yh675

CP626@CORNELL.EDU

SL2327@CORNELL.EDU

YH675@CORNELL.EDU

Abstract

We propose a novel model for time-series data using a Bayesian non-parametric framework. Training data is clustered into characteristic segments using a Dirichlet Process Gaussian Process. The Indian Buffet Process then is used to further learn a composition of kernel functions that models the characteristic segments. Finally, we propose a framework to improve accuracy of prediction for online inference and extrapolation of the time-series. We consider the improvement on scalability of the model during online inference due to evaluation of the clusters rather than the entire dataset.

1. Motivation

Time-series data in several domains is highly volatile and difficult to predict (for example, stock prices fluctuate wildly and are difficult to model accurately) due to their often non-stationary nature. Essentially, different locations in input space produce outputs described with variable functions. We propose an algorithm to model the non stationarity of time-series data by using a clustering method to segment time series data into different characteristic segments that correspond to the varying functions (with respect to input space) that model the outputs. For each cluster we will utilize the Indian Buffet Process (IBP) to learn a composition of kernel functions that models the underlying structures. Modeling the data in kernel function space allows for the flexibility of a non-parametric model while also capturing uncertainty in a probabilistic manner. Our approach is well-suited for non-stationary time-series data because it captures the characteristic segments of time-series data in a formal probabilistic model. Our model also allows for kernel composition learning in parallel which makes it scalable. In addition, given a model of time-series data, the ability for online prediction is desired. Generative time-series models have applications in a broad range of appli-

cations such as stock market prediction, earthquake detection, and weather forecasting. As more data is observed, the uncertainty in online predictions should decrease.

2. Related Work (Relevant Papers)

Previous work has modeled time-series data using purely non-parametric models (Hensman et al., 2014) and a combination of parametric and non-parametric models (Duvenaud et al., 2013). Parametric models generally require less data to learn but do not have the inherent flexibility of non-parametric models. Data-driven approaches have modeled time-series data with Gaussian Processes (GP) and Dirichlet Process Gaussian Process (DPGP) (Ross & Dy, 2013). These approaches have had success in clustering multiple trajectories of time-series data, but do not identify characteristic segments of a single time-series trajectory, which would potentially allow for higher accuracy during online inference. Other popular approaches to modeling time-series data find change points, which correspond to boundaries between characteristic segments, with a run-length variable and models the characteristic segments between change points accordingly (Saatci et al., 2010). A drawback of this type of approach is that the prior over the distribution of change points is not appropriate for modeling real data. A more statistically rigorous approach is desired. There has been application of the IBP to clustering multiple time-series trajectories into GP kernels (Tong & Choi, 2017). However, this approach does not find correlations among characteristic segments for single trajectory time-series data. Modeling with GPs can cause scalability problems due to the complexity of inverting the kernel matrix; the community has developed approximation methods that exploit kernel matrix structure for fast inference (Wilson & Nickisch, 2015).

Online learning with GPs expects Bayesian updates of the mean function, kernel, and posterior. (Csato & Opper, 2002) showed that by unfolding the recursion steps in the update rules we arrive at the parametrization for the approximate posterior GP as a function of the initial kernel and the likelihoods, which is a sequential propagation of an approximate Gaussian distribution. However, the parameter space will grow as more data is observed which

can cause intractability with massive datasets. To ease this problem, sparseness is introduced into the parameter space to approximate the posterior. However, the scalability still bottlenecks with big data applications.

3. Technical Approach

Our novel algorithm will (1) identify characteristic segments of time-series data using DPGP clustering, (2) perform kernel learning of the characteristic sections using the IBP and GPs, and (3) perform online inference.

3.1. Offline Training

3.1.1. DPGP CLUSTERING (COMPLETION BY 11/10)

We propose a method to identify characteristic sections of time-series data by using DPGP. The number of clusters or characteristic segments are drawn from a DP. A GP is fit over each cluster and hyperparameters are optimized using marginal likelihood. Gibbs sampling is employed to iteratively sample the clusters until satisfactory performance is achieved. Possible kernel functions to use include a Deep Kernel Learning (DKL) function (to exploit advantages of a deep network architecture and the generalization ability of a Gaussian Process (GP)) or a Spectral Mixture Kernel (due to its ability to fit periodic data well). The points at the ends of the characteristic sections are intuitively the change points and it would be an interesting exercise to correlate those change points with physical events.

3.1.2. IBP KERNEL LEARNING (COMPLETION BY 11/21)

After clustering data into characteristic segments, we will cluster each characteristic segment to kernels via the IBP. This second clustering process will find different compositions of kernels for the characteristic segments. The advantage of IBP interpretability from the compositions of kernels. For example, one seasonal segment would be assigned to both periodic and RBF kernel functions. Then, hyperparameters for the composition of kernels will be found by maximizing the marginal likelihood. We will also explore using DKL as another approach to finding the kernel functions of characteristic segments.

3.2. Online Inference (Completion by 11/27)

As data is observed online it is either clustered with an existing characteristic segment which limits the number of new parameters introduced, or a new cluster is formed. New clusters, which form new characteristic segments, are formed when the posterior of the new data belonging to an existing characteristic segment falls below a predefined threshold.

4. Midterm Milestone

We plan to have our algorithm fine-tuned for offline training and also be able to perform short term online inference. We will have developed the IBP process to be able to learn compositions of kernels from the DPGP clusters of characteristic segments. Relevant feedback from the midterm evaluations will be considered.

5. Evaluation

We provide criteria of what we would consider a perfect project and the minimum we expect to accomplish for the project to be considered a success.

5.1. Perfect Project

Be able to cluster characteristic segments of time-series data with DPGP and use IBP to find compositions of kernels describing the characteristic segments. We will have integrated DKL into the clustering algorithm for its flexibility to model variable data. Our generative model will be able to perform accurate long term online inference by exploiting the learned kernels of characteristic segments. The model will be scalable to large datasets.

5.2. Minimum

Develop a novel algorithm to find characteristic segments of time-series data with DPGP and use IBP to find compositions of kernels for the characteristic segments. Perform short-term online inference. This algorithm will be able to capture characteristic segments from synthetic data.

6. What Each Person Will Do

Chawannut will develop a model that uses the IBP to learn compositions of kernels for each characteristic segment. He will also implement approximation techniques for the scalability of the model with **Yutao**. **Serena** will develop the online inference section of the model and prepare the dataset. She will also investigate a DKL approach to discover expressive kernel functions for characteristic segments. **Yutao** will develop an algorithm that uses DPGP to cluster the time-series data into characteristic segments and potentially identify distributions over change points.

7. Proposed Dataset

We propose to use Bitcoin market price data which is publicly available ([Bit, 2017](#)). The data has clear characteristic segments and intuitive change points. The data dates back to 2014 with the resolution of daily ticks. We will also use synthetic data as a baseline.

References

- Bitcoin exchange rates. Quandl, 2017.
- Csato, Lehel and Oppel, Manfred. Sparse on-line gaussian processes. *Neural Computation*, 2002.
- Duvenaud, David, Lloyd, James, Grosse, Roger, Tenenbaum, Joshua, and Zoubin, Ghahramani. Structure discovery in nonparametric regression through compositional kernel search. In *Proceedings of the 30th International Conference on Machine Learning*. JMLR: W&CP, 2013.
- Hensman, James, Rattray, Magnus, and Lawrence, Neil D. Fast nonparametric clustering of structured time-series. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37:383–393, 2014.
- Ross, James C. and Dy, Jennifer G. Nonparametric mixture of gaussian process with constraints. In *Proceedings of the 30th International Conference on Machine Learning*. JMLR: W&CP, 2013.
- Saatci, Yunus, Turner, Ryan, and Rasmussen, Carl Edward. Gaussian process change point models. In *Proceedings of the 27th International Conference on Machine Learning*. JMLR, 2010.
- Tong, Anh and Choi, Jaesik. Discovering explainable latent covariance structure for multiple time series. 2017.
- Wilson, Andrew Gordon and Nickisch, Hannes. Kernel interpolation for scalable structured gaussian process (kiss-gp). In *Proceedings of 32nd International Conference on Machine Learning*. JMLR: W&CP, 2015.