Geophysical Research Abstracts
Vol. 12, EGU2010-9920, 2010
EGU General Assembly 2010

Variational Gaussian-process factor analysis for modeling spatio-temporal data

Alexander Ilin and Jaakko Luttinen
Aalto University, School of Science and Technology, Finland (alexander.ilin@tkk.fi)

article

amsmath bm graphicx

Variational Gaussian-process factor analysis for modeling spatio-temporal data

Alexander Ilin and Jaakko Luttinen
Adaptive Informatics Research Center
Aalto University, School of Science and Technology, Finland
firstname.lastname@tkk.fi

We present a new probabilistic model which can be used for studying spatio-temporal datasets [1]. The method is based on the factor analysis model $\mathbf{Y} = \mathbf{W}\mathbf{X} + noise = \sum_{d=1}^{D} \mathbf{w}_{:d}\mathbf{x}_{d:}^T + noise$, where $\mathbf{Y}$ is a matrix in which each row contains measurements of some quantity in one spatial location and each column corresponds to one time instance. Each vector $\mathbf{x}_{d:}$ (the d-th row of $\mathbf{X}$) represents the time series of one of the D factors whereas $\mathbf{w}_{:d}$ (the d-th column of $\mathbf{W}$) is a vector of loadings which are spatially distributed. Matrix $\mathbf{Y}$ may have missing values as the samples can be unevenly distributed in space and time.

The said model for $\mathbf{Y}$ yields standard principal component (or empirical orthogonal functions, EOF) analysis when both factors $\mathbf{x}_{d:}$ and noise are normally distributed and the noise variance is the same for each measurement. In our approach, we assume that factors $\mathbf{x}_{d:}$ and corresponding loadings $\mathbf{w}_{:d}$ have prominent structures such as, for example, slowness or periodicity for $\mathbf{x}_{d:}$ and spatial smoothness for $\mathbf{w}_{:d}$. We model such regularities using Gaussian processes (GPs), which is a flexible tool for smoothing and interpolating non-uniform data [2]. Applying the GP methodology directly to observations $\mathbf{Y}$ can be unfeasible in real-world problems because the computational complexity of the inference scales cubically w.r.t. data dimensionalities. Using separate GP models for $\mathbf{x}_{d:}$ and $\mathbf{w}_{:d}$ facilitates analysis of large spatio-temporal datasets because we perform GP modeling only either in the spatial or temporal domain at a time.

The model is identified using the framework of variational Bayesian learning. The true posterior distribution of the unknown variables $\mathbf{x}_{d:}$ and $\mathbf{w}_{:d}$ is approximated using a probability density function which has a tractable form. The parameters defining the GP priors are found by maximizing the lower bound of the respective likelihood. The computational complexity of GP modeling is reduced by using sparse approximations.

In the experimental part the model is used to compute reconstructions of historical sea surface temperatures. The global temperatures are computed for the period 1856-1991 using the data from the MOHSST5 dataset containing monthly sea surface temperature anomalies. We compare the proposed method with the state-of-the-art reconstruction methodology [3]. The advantage of the presented technique is that it incorporates the standard assumptions of the existing reconstruction systems in a single model, which allows for optimal use of all available data.

The proposed technique can be seen as a combination of such techniques as EOF analysis, kriging and temporal smoothing. The elegancy of the model is that it is completely symmetrical w.r.t. space and time. The model has good interpretability, which makes it easy to explore the results in the spatial and temporal domains and to set priors reflecting the modeling assumptions.

References

[1] J. Luttinen and A. Ilin. Variational Gaussian-process factor analysis for modeling spatio-temporal data, *NIPS 2009*.

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

[3] A. Kaplan, M. Cane, Y. Kushnir, A. Clement, M. Blumenthal and B. Rajagopalan. Analysis of global sea surface temperatures 1856–1991, *Journal of Geophysical Research*, 103:18567–18589, 1998.