Supervised classification of multidimensional and irregularly sampled signals
Alexandre Constantin, Mathieu Fauvel, Stéphane Girard, Serge Iovleff

To cite this version:
Alexandre Constantin, Mathieu Fauvel, Stéphane Girard, Serge Iovleff. Supervised classification of multidimensional and irregularly sampled signals. Statlearn 2019 - Workshop on Challenging problems in Statistical Learning, Apr 2019, Grenoble, France. pp.1. hal-02092347

HAL Id: hal-02092347
https://hal.archives-ouvertes.fr/hal-02092347
Submitted on 8 Apr 2019
Supervised classification of multidimensional and irregularly sampled signals.

Alexandre Constantin1, Mathieu Fauvel2, Stéphane Girard1 and Serge Iovleff3

1 Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, Grenoble, France
2 CESBIO, Université de Toulouse, CNES/CNRS/IRD/UPS/INRA, Toulouse, France
3 Laboratoire Paul Painlevé - Université Lille 1, CNRS, Inria, France

Introduction

Background:
Recent space missions, such as Copernicus Sentinel-2, provide high resolution Satellite Image Time Series (SITS) to study continental surfaces, with a very short revisit period (5 days for sentinel-2). In order to process such data, statistical models are regularly used [1, 2], which usually require a regular temporal sampling. However, for SITS, clouds and shadows (eg. figure from [3]), as well as the satellite orbit, an irregular temporal sampling is common.

Contribution:
A new statistical approach using Gaussian processes is proposed to classify irregularly sampled signals without temporal rescaling. Moreover, the model offers a theoretical framework to impute missing values such as cloudy pixels.

Model

Gaussian Processes (GP) model:
Let $S = \{(y_i, z_i)\}_{i=1}^n$ be a set of multidimensional and irregularly sampled signals. A signal $Y$ is modeled as a vector of $p$ independent random processes $\mathcal{T} \to \mathbb{R}^p$, with $\mathcal{T} = \{0, T\}$. The associated label is modeled by a discrete random variable $Z$ taking its values in $\{1, \ldots, C\}$. The model introduced here is based on two assumptions: 1) The coordinate processes $Y_{b,c} \ b \in \{1, \ldots, p\}$ of $Y$ are independent, 2) Each process $Y_i$ is, conditionally to $Z = c$, a Gaussian process. Then

$Y_i(t) | Z = c \sim \mathcal{GP}(m_{b,c}(t), K_{b,c}(t, s))$,

where $m_{b,c} : \mathcal{T} \to \mathbb{R}^p$ is a mean function, and $K_{b,c}$ a covariance kernel with hyperparameters $\theta_{b,c}$. For example $\theta_{b,c} = \{\gamma_{b,c}^2, \kappa_{b,c}, \sigma_{b,c}^2\}$ with

$K_{b,c}(t, s) = \gamma_{b,c}^2 \kappa(t, s) \delta_{b,c} + \sigma_{b,c}^2 \delta_{b,c}$

An irregularly sampled noisy signal $y_i$ is observed on $T_i$ time stamps $\{t_1, \ldots, t_{T_i}\} \in \mathcal{T}$ and its $b$th coordinate is represented by a vector in $\mathbb{R}^p$. We write $y_{i,b} = [Y_{i,b}(t_1), \ldots, Y_{i,b}(t_{T_i})]^T$, with

$y_{i,b} | Z_i = c \sim N_{T_i}(\mu_{i,b,c}, \Sigma_{i,b,c})$.

Estimation:
• $\alpha_{b,c}$ and $\theta_{b,c}$ are estimated by maximizing the log-likelihood,

$-\frac{1}{2} \sum_{i: |Z_i = c} \log \left| \Sigma_{i,b,c} \right| + (y_{i,b} - B_i \alpha_{b,c})^T \Sigma_{i,b,c}^{-1} (y_{i,b} - B_i \alpha_{b,c})$.

• $\alpha_{b,c}$ is given by an explicit formula, while $\theta_{b,c}$ is computed thanks to a gradient technique.

Classification and imputation of missing values

The assigned class is given by the MAP rule from the posterior probability

$P(Z = c | y_i) = \frac{\pi_c \prod_{b,c} P_{b,c}(y_{i,b} | B_i \alpha_{b,c} \Sigma_{(\theta_{b,c}))})}{\sum_{c'} \pi_{c'} \prod_{b,c} P_{b,c}(y_{i,b} | B_i \alpha_{b,c} \Sigma_{(\theta_{b,c}))})}$

When the class is known to be $c$, the missing value at $t^*$ is estimated through the computation of conditional expectation.

$\hat{Y}_{b,c}(t^*) = B(t^*) \alpha_{b,c} + K_{b,c}(t^*, t_1 : T)^T \Sigma(\theta_{b,c})^{-1} (y_{i,b} - B(t^*) \alpha_{b,c})$

$\text{var}(\hat{Y}_{b,c}(t^*)) = K_{b,c}(t^*, t^*) - K_{b,c}(t^*, t_1 : T)^T \Sigma(\theta_{b,c})^{-1} K_{b,c}(t_1 : T, t^*)$

We also generalized this imputation when the class is unknown.

Future work

We are now implementing the model for massive real data (Sentinel-2). We are also working on a new model when the bands are correlated.

This work is supported by the French National Research Agency in the framework of the Investissements d’Avenir program (ANR-15-IDEX-02) and by the Centre National d’Études Spatiales (CNES).

References: