Pattern Recognition Letters 38 (2014) 1–7

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

A nonlinear mixed-effects model for simultaneous smoothing and registration of functional data

Lars Lau Rakêt [a,*], Stefan Sommer [a], Bo Markussen [b]

[a] Department of Computer Science, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark
[b] Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 24 June 2013
Available online 30 October 2013

Communicated by Y. Liu

Keywords:
Data alignment
Functional mixed-effects model
Nonlinear mixed-effects model
Phase variation
Amplitude variation
Smoothing

ABSTRACT

We consider misaligned functional data, where data registration is necessary for proper statistical analysis. This paper proposes to treat misalignment as a nonlinear random effect, which makes simultaneous likelihood inference for horizontal and vertical effects possible. By simultaneously fitting the model and registering data, the proposed method estimates parameters and predicts random effects more precisely than conventional methods that register data in preprocessing. The ability of the model to estimate both hyperparameters and predict horizontal and vertical effects are illustrated on both simulated and real data.

1. Introduction

The current standard practice of analyzing functional data in a number of sequential steps is problematic. Analyses are often carried out by performing one or more independent preprocessing steps prior to the final statistical analysis (Ramsay and Silverman, 2005). Typical examples are data registration, pre-smoothing, and dimensionality reduction. Such preprocessing steps can cause problems since the final analysis does not take the resulting data modifications (and their related uncertainty) into account. In the worst case this may invalidate the conclusions of the final analysis.

This paper considers misaligned functional data, where proper registration is key to analyzing the data. Treating data registration as a preprocessing step can cause problems. In particular, noisy observations can skew registration results such that noise rather than signal is aligned. Since this type of overfitting happens prior to the statistical analysis, it will lead to both wrongly predicted warps and underestimation of the noise variance. To deal with these issues we propose to simultaneously do likelihood-based smoothing and data registration in a general class of nonlinear functional mixed-effects models. By computing both registration and smoothing at the same time, we will get the optimal registration given the prediction of the functional mixed-effects and vice versa.

The mixed effects are assumed to be observations of Gaussian processes, and the resulting calculations are carried out by iteratively linearizing the model and estimating parameters from the resulting likelihood function. In addition to allowing estimation of the optimal combination of smoothing and registration, all parameters can be estimated by maximum-likelihood estimation. This contrasts most previous works on simultaneous smoothing and registration (see e.g. Lord et al. (2007) and Kneip and Ramsay (2008)) where parameters have to be adjusted (semi-)manually. Some notable exceptions are Rønn (2001); Gervini and Gasser (2005); and Rønn and Skovgaard (2009) who presents methods for doing full likelihood inference for time-transformed curves, and Allassonnière et al. (2007) who derive a rigorous Bayesian framework for estimating data deformation and related parameters. In contrast to the mentioned works, the model we present seeks to align fixed effects, but allows for serially correlated effects that cannot be matched across functional samples. Since much functional data contains serially correlated noise, e.g. from the measuring device or individual sample differences, a model that allows the separation of such amplitude variations from the phase variation is a considerable step forward.

It is worth noting the differences with pair-wise data registration as is often employed in for example medical imaging. Instead of the common approach of choosing parameters of the registration model either by heuristic arguments or by cross-validation, incorporating the entire dataset or population in the analysis allows parameters to be estimated by maximum-likelihood inference. In addition, instead of searching for a similarity measure

* Corresponding author. Tel.: +45 353 21400.
E-mail addresses: larslau@diku.dk (L.L. Rakêt), sommer@diku.dk (S. Sommer), bomar@life.ku.dk (B. Markussen).

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.10.018

that is invariant to certain types of serially correlated effects, e.g. mutual information (Viola and Wells, 1995), the explicit modeling of the serially correlated effects removes the need for invariance in the similarity measure.

The proposed methods are illustrated and compared to conventional preprocessing alignment on simulated dataset, and a general model for alignment is proposed and evaluated on four real datasets.

2. Motivation and preliminaries

Two of the major challenges when analyzing functional data are modeling of individual sample effects and aligning of functional samples. Fig. 1 illustrates these effects on their own, and in combination, on a one-dimensional functional dataset.

In order to handle individual variation (corresponding to the situation in Fig. 1(a)), one can consider a linear functional mixed-effects model where the kth observation point of functional sample i from the dataset $\boldsymbol{y}$ is assumed to be generated as follows

$$y_i(t_k) = \theta(t_k) + x_i(t_k) + \varepsilon_{ik}, \tag{1}$$

where θ is a fixed effect, x_i is a zero-mean Gaussian process with covariance function $\sigma^2\mathcal{S}$, and ε_{ik} is independent identically distributed Gaussian noise with variance σ^2. Inference in this class of models has been considered in numerous works (Guo, 2002).

In contrast to the vertical variation due to individual sample differences one may encounter horizontal variation due to non-aligned samples (Fig. 1(b)). To align samples, one wishes to estimate so-called *warping functions* v that model the horizontal variation. Similarly to the vertical variation, one may consider the following functional mixed-effects model for this setup

$$y_i(t_k) = \theta(v(t_k, \boldsymbol{w}_i)) + \varepsilon_{ik}, \tag{2}$$

where θ and ε_{ik} are as in (1), and v is a warping function depending on $\boldsymbol{w}_i$ that is a vector of Gaussian parameters with covariance matrix C_0. This model can be considered a nonlinear mixed-effects model, and many known registration algorithms can be thought of as methods for predicting the warping parameters in the model (2), with a known fixed effect θ.

The model (2) has been considered in a statistical setting by Rønn (2001); Gervini and Gasser (2005); and Rønn and Skovgaard (2009), who all consider the problem in a nonparametric maximum likelihood setting. An alternative view is taken in shape analysis, where the interest is on the common shape θ, while the warping functions are considered nuisance parameters, and data is generally considered free of observation noise. From this viewpoint Kurtek et al. (2011) and Srivastava et al. (2011) have recently proposed an estimation procedure for θ based on the Fisher–Rao metric, that is invariant to diffeomorphic data warping. The mean shape is subsequently used for estimating the warping functions and aligning data. This approach produces state-of-the-art results on numerous examples, but is not generally applicable to all types of data, since the invariance to diffeomorphic warping may lead to overfitting when significant noise is present.

In practice, data often exhibit both vertical and horizontal variation. Fig. 1(c) shows alignment variations of the fixed effect with added serially correlated effects, i.e. a combination of the models (1) and (2)

$$y_i(t_k) = \theta(v(t_k, \boldsymbol{w}_i)) + x_i(t_k) + \varepsilon_{ik}. \tag{3}$$

This type of model describe the fixed effect as a deformation of θ and allows a serially correlated effect x_i that follows the coordinate system of the observation. For some examples, it may be natural to consider the correlated effects x_i in the coordinate system of the fixed effect θ. That model will not be considered here, but inference may be done completely analogous to the procedure described for model (3).

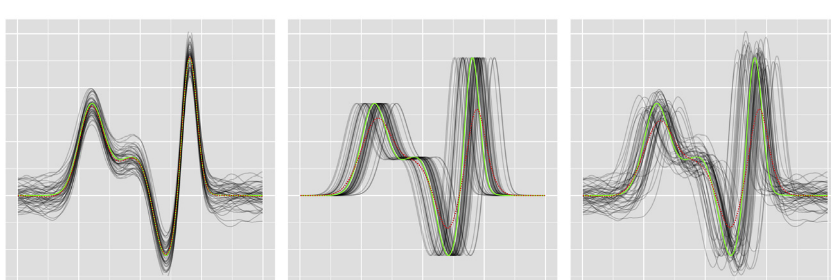
Data modeling following the lines of model (3) have received little attention. One notable exception is the paper by Bigot and Charlier (2011) who consider the sample Fréchet mean as an estimator for θ in the model (3) where the effect x_i also undergo warping by v, and give conditions under which the estimator is consistent. They do however not consider parameter estimation and prediction of random effects. In another related work, Elmi et al. (2011) derive a B-spline based nonlinear mixed-effects model in a maximum likelihood setting. The model allows incorporation of data registration, and is applied to labor curve data, where amplitude variation is modeled parametrically, with random additive and multiplicative effects. Another application of this type of model is considered by Chambolle and Pock (2011) in the setting of motion estimation in image sequences. They propose to include a spatially correlated effect that plays the role of lighting differences between the images in question. Their approach, however, does not take the uncertainty related to the prediction of the spatially correlated effect into account in the estimation of the warp, and do not consider the question of parameter estimation.

In the following we will derive inference methodology for the model (3). In contrast to conventional preprocessing approaches that register raw data, the proposed methods can separate horizontal and vertical variation, and allows for maximum-likelihood estimation of all hyperparameters.

3. Estimation

Consider model (3), where the functional data is defined on a domain $\mathcal{T} \subseteq \mathbb{R}$, with m vectorized samples $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$, each of which consists of n points.

The estimation procedures consists of interleaved steps of estimating (a) the fixed effect and the warps; and (b) the parameters of the model and the serially correlated effects. In order to do



(a) Individual variations (b) Alignment variations (c) Alignment variations plus individual variations

Fig. 1. Different types of variation in a one-dimensional functional dataset. The true underlying curve is shown in green, the average curve is shown in dashed red. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)