Curve Clustering via the Split Learning of Mixtures of Gaussian Processes

Zhe Qiang, Jiahui Luo, Jinwen Ma
Department of Information Science,
School of Mathematical Sciences & LMAM, Peking University,
Beijing, 100871, P.R. China
Email: jwma@math.pku.edu.cn

***Abstract*—Data in various research fields can be gathered as repeated measure curves. Although they consist of finite points, it is usually valuable to consider them as sample curves of stochastic processes so that curve clustering is necessary for the modelling and analysis of these latent stochastic processes. In this paper, we model these curves through a hierachical mixture of Gaussian processes and propose a split EM algorithm to learn the parameters of the mixture of Gaussian processes with automated model selection and thus to cluster these curves according to the Gaussian processes. Specifically, during each iteration of the proposed split EM algorithm, one Gaussian process is selected with the highest acceptance probability for splitting and then it is split into two Gaussian processes whose first two moments keep the same as those of it. It is demonstrated by the experiments on both synthetic and real datasets that our proposed split EM is robust and effective for curve clustering and even outperform the conventional EM algorithms.**

Keywords:Curve clustering, Mixture of Gaussian processes, EM algorithm, Parameter estimation, Model selection

I. Introduction

Curve Clustering is an important but very challenging problem [1] [2]. When handling repeated measure data, we often need to register or align the corresponding curves. By the method of curve clustering, we can step over curve alignment and classify the curves according to their variation directly.

However, there are two difficulties on curve clustering: the first is how to transform a vector to a piece of stochastic process–although data are gathered as finite points, they have to be considered as a time series or a sample curve of stochastic process; the second is how to model the heterogeneity—Data may be collected from different sources, such as different subject or region. In fact, the hierachical mixture of Gaussian process functional regressions (Mix-GPFR) [3] [4] [5] [6] can be used to overcome these two difficulties. Firstly, a GPFR can effectively model each sample or response curve with both the covariance parameters and the mean function which is estimated by a linear combination of some given B-spline basis functions. Secondly, the mixture model can be used to model heterogeneity of curves from different sources. Recently, our lab has made a series of progresses on the parameter learning and model selection of the mixture of Gaussian process functional regressions [7] [8] [9] [10] [11].

In this way, curve clustering is transformed into the modelling or learning of the Mix-GPFR model with a given set of response curves. The possible approach to solving this mixture learning problem is the EM algorithm. Although many efforts have been made to construct an effective and efficient EM algorithm for the mixtures of Gaussian processes, this task is still difficult. Moreover, the EM algorithm is sensitive to the parameter initialization and may converge to a local maxima of the log-likelihood function. On the other hand, the EM algorithm requires the correct number of components in the mixture; otherwise it leads to a wrong result. So, it is critical for the EM algorithm to select the correct number of components in the mixture on a given dataset. This is a well-known model selection problem for the mixture modeling and has been investigated in many ways (e.g., [12], [13], [14], [15], [16]).

In order to solve the curve clustering problem, we try to propose a split EM algorithm for the Mix-GPFR model which starts with a single Gaussian process component standing for all of the curves and then split one Gaussian process with the highest acceptance probability for splitting into two Gaussian processes whose first two moments keep the same as those of the original one in each of the following iterations until certain model selection criterion is satisfied. The key to the success is the possible split criterion. In fact, various split criterions have already been developed for the Gaussian mixture model on the same model selection problem, but they are not so suitable for our case as Gaussian distribution and Gaussian process have quite different structures. Fortunately, our recently proposed automatic split and merge scheme for Gaussian processes in the mixture model under the MCMC framework [17] provides an effective split criterion to the split EM algorithm with automated model selection. That is, by keeping the first two moments of those GPs unchanged before and after split, we can derive a series of split formula and calculate the acceptance probability for splitting under the Bayesian condition. Moreover, we adopt BIC for model selection in the design of the split EM algorithm.

Our paper is organized as follows. We introduce the hierachical MGP or mix-GPRF model in Section 2. Section 3 presents the split EM algorithm for the mix-GPRF model. Specifically, Section 3.1 is devoted to the presentation of split formula in our previous work. section 3.2 presents the detail of the split EM algorithm. Section 4 and 5 contain the experimental results on the synthetic and real datasets, respectively. Finally, we conclude briefly in Section 5.

Authorized licensed use limited to: Peking University. Downloaded on June 03,2022 at 10:03:45 UTC from IEEE Xplore. Restrictions apply.

II. The Hierarchical Mixture of Gaussian Processes

We adopt the hierachical mixture of Gaussian processes which was firstly proposed in [4]. The m-th response curve is considered as the k-th class stochastic process with a noise corrosion: $y_m = f_k(x_m) + \epsilon_k$. For the heterogeneity, a set of hidden variable $z_m = 1; \cdots ; K$ for each curves that describes which class the curve belong to is set to model K Gaussian processes in the mixture. That is, for curves in the same class, a common series of Gaussian process parameters are shared. That is to say:

$$\begin{aligned} z_m &= k \\ \boldsymbol{y}_m &\sim \mathcal{N}(\boldsymbol{\mu}_k(\boldsymbol{x}_m); \boldsymbol{\Sigma}_k(\boldsymbol{x}_m)) \end{aligned} \tag{1}$$

here, the vector pair $\mathcal{D}_m = (\boldsymbol{x}_m; \boldsymbol{y}_m)$ represents the m-th curve, and $\boldsymbol{x}_m = \{x_{m,i}\}_{i=1}^{N_m}; \boldsymbol{y}_m = \{y_{m,i}\}_{i=1}^{N_m}$ are points on the mth curve.

For the k-th Gaussian process, the mean function is assumed as a linear combination of B-spline basis functions [3]:

$$\boldsymbol{\mu}_k(x_m) = \boldsymbol{\Phi}(x_m)\boldsymbol{B}_k; \tag{2}$$

where $\boldsymbol{\Phi}(x) = (\boldsymbol{\Phi}_1(x); \cdots ; \boldsymbol{\Phi}_D(x))$ is a set of D B-spline basis functions, $\boldsymbol{\Phi}(x_m)$ is an $N_m \times D$ matrix, $B_k = (B_k^1; \cdots ; B_k^D)^T$ is $D \times 1$ unknown B-spline coefficient matrix.

And the convariance function is:

$$\Sigma_k(x_m) = v_k \, \mathrm{e} \, \mathrm{p}\left(-\frac{w_k}{2}(x_{m,i} - x_{m,j})^2\right) + \ _{i,j} \ ^2_k \tag{3}$$

Therefore, the parameters for each GP class are $(\ _k; \boldsymbol{\theta}_k)$, $\boldsymbol{\theta}_k = (w_k; v_k; \ ^2_k)$. And all the parameters are $(\boldsymbol{\Pi}; \boldsymbol{\Theta})$, where $\boldsymbol{\Pi} = \{\ _k\}_{k=1}^K; \boldsymbol{\Theta} = \{\boldsymbol{\theta}_k\}_{k=1}^K$.

After learning all the parameters with the EM algorithm, the prediction of the m-th curve obeys the following formulas [18]:

$$\begin{aligned} E[f_m(x^*)|\mathcal{D}_m] &= \ ^T(x^*)\Sigma^{-1}(\boldsymbol{X}_m; \boldsymbol{X}_m; \boldsymbol{\theta}_k)\boldsymbol{Y}_m; \\ Var[f_m(x^*|\mathcal{D}_m)] &= C(x^*; x^*) - \ ^T(x^*)\Sigma^{-1}(\boldsymbol{X}_m; \boldsymbol{X}_m; \boldsymbol{\theta}_k) \ (x^*); \end{aligned} \tag{4}$$

III. Split EM Algorithm

We begin with a brief introduction to the split formulas and the probability to accept a proposed move according to our previous work [17]. In the following subsection, we will try to utilize it in the design of the split EM algorithm.

A. The Split Formulas

The formulas given here are derived by the Bayesian inference, so all the parameters have their own priors:

$$w_k \sim I\Gamma(\frac{1}{2}; \frac{1}{2}); v_k \sim \mathcal{LN}(-1; 1^2); \ ^2_k \sim \mathcal{LN}(-3; 3^2); k = 1; \cdots ; K$$

where $I\Gamma$ denotes the inverse gamma distribution, and

$$(\ _1; \cdots ; \ _K) \sim Dir(1; \cdots ; 1):$$

During the split operation, the first two moments must remain constant:

$$\begin{aligned} \ _{k*} &= \ _{k_1} + \ _{k_2} \\ \ _{k*}\Sigma_{k*} &= \ _{k_1}\Sigma_{k_1} + \ _{k_2}\Sigma_{k_2} \end{aligned}$$

replacing Σ_k with eq.3 and transforming the nonlinear relationship of convariance parameters into linear relationship by Taylor expansion, we have the detailed balance framework:

$$\ _{k^*} = \ _{k_1} + \ _{k_2} \tag{5a}$$

$$\ _{k^*} \ ^2_{k^*} = \ _{k_1} \ ^2_{k_1} + \ _{k_2} \ ^2_{k_2} \tag{5b}$$

$$\ _{k^*} v_{k^*} = \ _{k_1} v_{k_1} + \ _{k_2} v_{k_2} \tag{5c}$$

$$\ _{k^*} v_{k^*} w_{k^*} = \ _{k_1} v_{k_1} w_{k_1} + \ _{k_2} v_{k_2} w_{k_2} \tag{5d}$$

For a split move, according to the reversible jump theory [5], 4 dimensional random vector $\boldsymbol{u}$ should be generated, where $u_i \sim Beta(2; 2)$, so that the dimension-matching requirement in the Reversible Jump MCMC framework can be satisfied. Thus, by combining the detailed balance framework and reversible jump theory, we have the following split formulas:

$$\ _{k_1} = u_1 \ _{k^*}; \quad \ _{k_2} = (1 - u_1) \ _{k^*}; u_1 \in (0; 1) \tag{6a}$$

$$\ ^2_{k_1} = u_2 \ ^2_{k^*} \frac{\ _{k^*}}{\ _{k_1}}; \quad \ ^2_{k_2} = (1 - u_2) \ ^2_{k^*} \frac{\ _{k^*}}{\ _{k_2}}; u_2 \in (0; 1) \tag{6b}$$

$$v_{k_1} = u_3 v_{k^*} \frac{\ _{k^*}}{\ _{k_1}}; \quad v_{k_2} = (1 - u_3) v_0^{k^*} \frac{\ _{k^*}}{\ _{k_2}}; u_3 \in (0; 1) \tag{6c}$$

$$w_{k_1} = \frac{1 - u_4}{u_3} w_{k^*}; \quad w_{k_2} = \frac{u_4}{1 - u_3} w_{k^*}; u_4 \in (0; 1) \tag{6d}$$

According to the acceptance probability formula of the reversible jump move [5], the acceptance ratio for a split move is $\min(1; A)$, where A is given by

$$\begin{aligned} A = \prod_{m=1}^{M} \frac{l(\boldsymbol{Y}_m|\boldsymbol{\theta}_{k+1})}{l(\boldsymbol{Y}|\boldsymbol{\theta}_k)} \times \frac{d_{k+1}}{b_k} \times \underbrace{k \frac{1}{6u_1(1-u_1)} \ _{k^*}}_{\pi} \\ \times \underbrace{\frac{1}{3\sqrt{2}} \frac{\ ^2_k}{\ ^2_{k_1} \ ^2_{k_2}} \times \frac{1}{6u_2(1-u_2)} \ ^2_{k^*} \frac{1}{u_1(1-u_1)}}_{\sigma^2} \\ \underbrace{\mathrm{e} \, \mathrm{p}\left(-\frac{(\ln \ ^2_{k_1} + 3)^2}{2 * 3^2} - \frac{(\ln \ ^2_{k_2} + 3)^2}{2 * 3^2} + \frac{(\ln \ ^2_{k^*} + 3)^2}{2 * 3^2}\right)}_{\sigma^2} \\ \times \underbrace{\frac{1}{\sqrt{2}} \frac{v_{k^*}}{v_{k_1} v_{k_2}} \frac{1}{6u_3(1-u_3)} v_{k^*} \frac{1}{u_1(1-u_1)}}_{v} \\ \underbrace{\mathrm{e} \, \mathrm{p}\left(-\frac{(\ln v_{k_1} + 1)^2}{2} - \frac{(\ln v_{k_2} + 1)^2}{2} + \frac{(\ln v_{k^*} + 1)^2}{2}\right)}_{v} \\ \underbrace{(\frac{1}{2})^{1/2} \left(\frac{w_{k_1} w_{k_2}}{w_{k^*}}\right)^{-3/2} \frac{1}{6u_4(1-u_4)} \frac{w_{k^*}}{u_3(1-u_3)}}_{w} \\ \underbrace{\mathrm{e} \, \mathrm{p}\left(-1{=}2\left(\frac{1}{w_{k_1}} + \frac{1}{w_{k_2}} - \frac{1}{w_{k^*}}\right)\right)}_{w} \end{aligned} \tag{7}$$

Authorized licensed use limited to: Peking University. Downloaded on June 03,2022 at 10:03:45 UTC from IEEE Xplore. Restrictions apply.

where $d_1 = 0; b_{k_{\max}} = 0; b_k = d_k = 0{:}5; \forall k = 2; \cdots; k_{\max} - 1$, $k_{\max}$ is the maximum component number that we set according to each individual case.

B. The Framework of the Split EM Algorithm

A full algorithmic description of the split EM algorithm is shown in Algorithm 1. It should be noted that our proposed algorithm consists of two layers of iterations when it decides which Gaussian process will be split: it runs to split all the K components and to sample $\boldsymbol{u}$ so that we can choose the one with the highest A (A in acceptance probability, that is calculated by Eq.(7). In case of choosing a larger component number with less increasing of the likelihood, we use BIC as a stop criterion. What's more, each curve should be minus the k-th mean function that this curve belong to, since our split criterion is based on the assumption of zero mean.

Algorithm 1 The split EM algorithm

Initialization: Start with one Gaussian process.
initialize Θ and Π according to their priors;
Execute the EM algorithm update all the paramters Θ , Π and B;
Split = false
repeat
  $y'_m = y_m - B_k * \Phi_m; m = 1; \cdots; K$, if $z_m = k$
  Use y'_m to execute **the split EM algorithm:**
  for $k = 1$ **to** K **do**
    for Iter $= 1$ **to** Iter_num **do**
      Generate a random vector $\boldsymbol{u}$
      Split k-th component according to Eq.(1)
      Calculate the acceptance ratio $A(k; Iter)$
    end for
    Set $Iter_* = \arg\max_{Iter=1\cdots Iter_num} A(k; Iter)$
    Ret $A(k) = A(k; Iter_*)$
    Record the $Iter_*$-th parameters
  end for
  Set $k_* = \arg\max_{k=1\cdots K} A(k)$
  Set $A = A(k_*)$
  Record the k_* parameters
  Execute the EM algorithm update all the parameters as Θ_{new} ,Π_{new} and B_{new}
  Calculate BIC between $\Theta_{new}; \Pi_{new}; B_{new}$ and $\Theta; \Pi; B$
  if BIC<0 **then**
    Split = true
    Accept the new parameters as Θ, Π and B
    Delete empty components
  else
    Split = false
    Remain the old parameters
  end if
until Split== false

IV. Experimental Results on a Synthetic Dataset

We firstly conduct an experiment on a typical synthetic dataset given in [7] to test the performance of our split EM algorithm. This synthetic dataset consists of three Gaussian processes with different parameters. The three mean functions are: $_1 = \frac{1}{2}\sin\left(-\frac{(x-4)^2}{8}\right) + 3$, $_2 = -\frac{3}{\sqrt{2\pi}}\mathrm{e}\ \mathrm{p}\left(-\frac{(x-4)^2}{8}\right) + 3{:}7$, $_3 = -\frac{1}{2}\arctan\left(\frac{x}{2} - 2\right) + 3$, and the convariance parameters of eq.(3) are listed in TABLE.I.

TABLE I
CONVARIANCE PARAMETERS

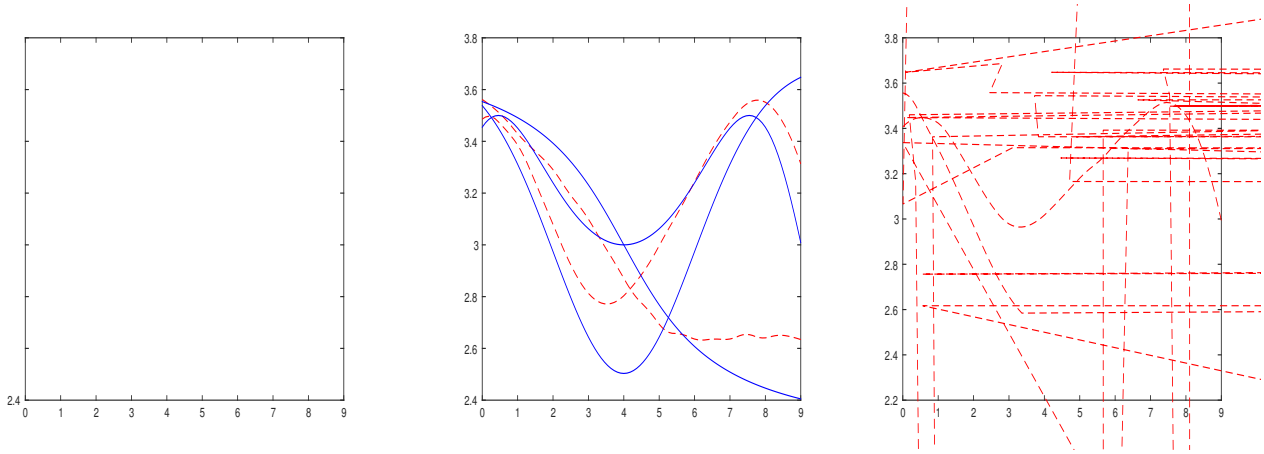
$\sqrt{v}$	$\sqrt{w}$	σ
0.6325	1	0.0632
0.4472	0.7071	0.0632
0.3162	0.4472	0.0632

We generate 300 sample curves from this model(each GP with 100 curves). For time-scale, we adopt 101 equally spaced points in $[0; 9]$, and choose 51 points randomly from those points as training data. For prediction, we generate 600 curves with each containing 150 points, and choose 40 points randomly as known data points on each curves, the rest as test data points.

As shown in Fig.1, we compare the real mean functions and B-spline fitted mean functions during the iterations of our split EM algorithm. In the one or 2 classes figure, B-spline fitted mean functions are in the middle of the true mean functions, and 3 classes B-spline fitted mean functions match the true mean-functions quite well.

With the same training and test dataset, we implement our split EM algorithm for (referred to as split-mix-GPFR) in comparison with the same split EM algorithm for mixtures of Gaussian processes (referred to as split-mix-GP), the conventional EM algorithm for mixtures of Gaussian process functional regressions (referred to as mix-GPFR) [3] and the conventional EM algorithm for mixtures of Gaussian processes (referred to as mix-GP) [4]). Each algorithm is repeatedly implemented 50 times and the experimental results are listed in TABLE.II. The mean of RMSE stands for the average RMSE on 50 trials amd the var of RMSE is the standard variance of RMSE on 50 trials. The times of $K = 3$ appears during 50 trials is the frequency of $K = 3$ and the misclassification rate is the average misclassification rate. It is remarkable that the RMSEs of the two algorithms for mixtures of GPFRs is smaller than those of the two algorithms for mixtures of GPs. Thus, GPFR can fit the data more accurately. This result shows that the GPFR model can improve the predict accuracy. Although the RMSEs of the two split EM algorithms are quite close to the two conventional EM algorithms, their RMSE variances are much smaller, which proves the robustness of our split EM algorithm. What's more, both of the two split EM algorithms can run with all the right component number–3. However, 2 times of model selection or component number error appears for mix-GPFR, 43 times of model selection error appears in mix-GP during 50 trials. However, the average misclassification rate can be reduced by both the functional regression model and the split EM mechanism.

Authorized licensed use limited to: Peking University. Downloaded on June 03,2022 at 10:03:45 UTC from IEEE Xplore. Restrictions apply.



(a) The 1st iteration (b) The 2sd iteration (c) The 3rd iteration

Fig. 1. The real mean function and B-spline fitted mean function during each iteration, where the blue line stands for the real mean function, while the red line stands fo the B-spline fitted mean function obtained in the split EM algorithm.

In addition to the prediction, we also list the average RMSEs of the convariance parameters in TABLE.III, in which the second row shows the real values of these parameters, while the rest rows show the learning results of mix-GPFR and split-mix-GPFR, respectively. The relative error is calculated by $\frac{|\hat{\theta}-\theta|}{|\theta|}$ where and ˆ are the real and estimated parameters. The var of error is the variance of an estimate parameter on 50 trials. By contrasting it is found that the relative errors of split-mix-GPFR are slightly smaller than those of mix-GPFR except $\sqrt{w_2}$ which is remarkable smaller. However, the variances of the estimate parameters of split-mix-GPFR are notably smaller than those of mix-GPFR, which further demonstrates the robustness of our split mechanism.

TABLE II
THE EXPERIMENTAL RESULTS OF 4 ALGORITHMS OVER 50 TRIALS

	mean of RMSE	var of RMSE
mix-GP	0.0850	2.9e-6
split-mix-GP	0.0877	3.1e-10
mix-GPFR	0.0735	9.5e-8
split-mix-GPFR	0.0734↓	3.5e-13↓
	the times of $K = 3$ appears during 50 repeat trials	misclassification rate
mix-GP	4/50	48.27%
split-mix-GP	50/50	5.67%
mix-GPFR	48/50	3.18%
split-mix-GPFR	50/50↑	0.33%↓

TABLE III
THE AVERAGE RMSES OF CONVARIANCE PARAMETERS OVER 50 TRIALS

	$\sqrt{v_1}$	$\sqrt{w_1}$	σ_1
real value	0.6325	1	0.0632
mix-GPFR	0.6271	0.9919	0.0631
relative error	0.84%	0.81%	0.31%
var	1.9e-5↑	1.3e-5↑	3.7e-8↑
split-mix-GPFR	0.6275	0.9927	0.0631
relative error	0.78%	0.73%	0.25%
var	3.0e-8↓	2.7e-9↓	3.9e-12↓
	$\sqrt{v_2}$	$\sqrt{w_2}$	σ_2
real value	0.4472	0.7071	0.0632
mix-GPFR	0.4467	0.7865	0.0634
relative error	0.1%	11.23%	0.23%
var	0.0052↑	0.2181↑	3.2e-6↑
split-mix-GPFR	0.4521	0.7014	0.0636
relative error	1.1%	0.81%	0.6%
var	5.3e-8↓	7.3e-8↓	3.4e-12↓
	$\sqrt{v_3}$	$\sqrt{w_3}$	σ_3
real value	0.3162	0.4472	0.0632
mix-GPFR	0.3411	0.4544	0.0634
relative error	7.87%	1.62%	0.31%
var	0.002↑	0.002↑	9.1e-9↑
split-mix-GPFR	0.3268	0.4414	0.0634
relative error	3.33%	0.27%	1.3%
var	4.4e-9↓	2.2e-10↓	1.5e-12↓

V. EXPERIMENTAL RESULTS ON A REAL DATASET

We further test our spit EM algorithm on a real dataset: Berkeley Growth Study data. It contains the heights of 36 males and 54 females from 1 age to 18 age. We want to cluster the total population according to the velocity and acceleration of height growth.

Firstly, these velocity and acceleration data are smoothed by a classic smoothing procedure in [19]. Secondly, let each point

Authorized licensed use limited to: Peking University. Downloaded on June 03,2022 at 10:03:45 UTC from IEEE Xplore. Restrictions apply.

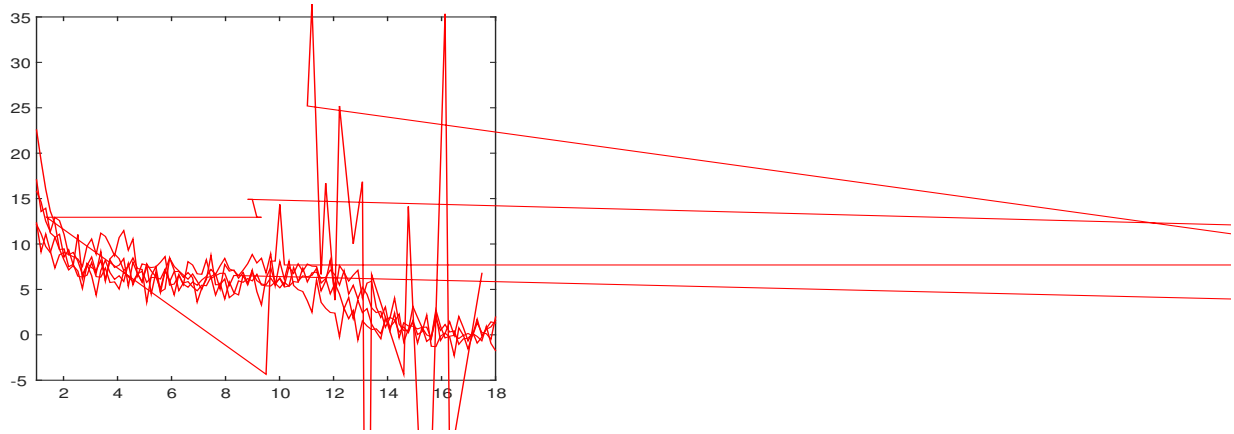
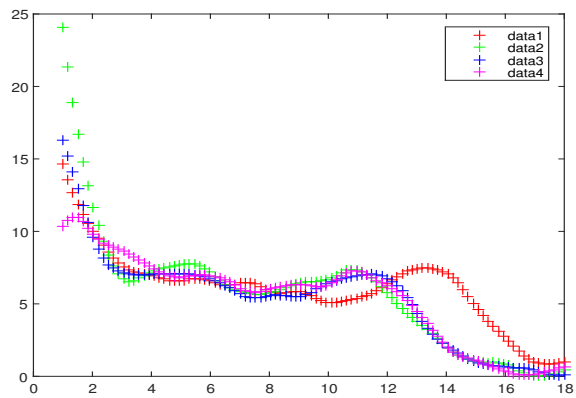
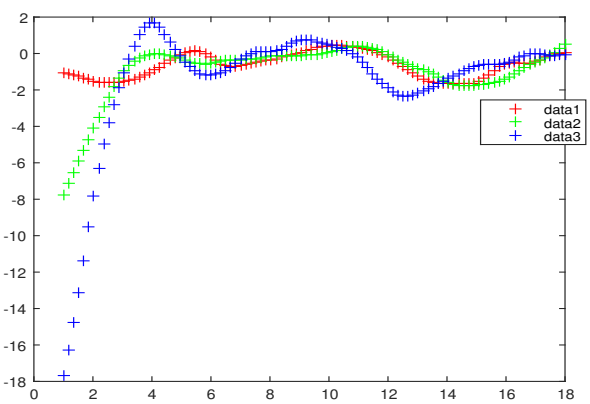


Fig. 2. The smooth and noisy velocity and acceleration curves. Red curves stand for female, green stand for male.

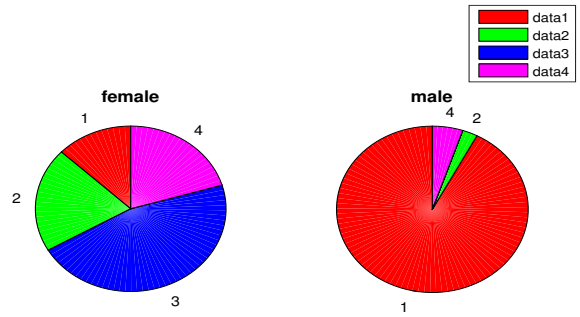


(a) The velocity mean functions clustered into 4 classes.



(b) acceleration mean functions clustered into 3 classes.

Fig. 3. The clustering results of the velocity and acceleration curves, with the mean function of each class being plot.



(a) The pie of velocity

(b) The pie of acceleration

Fig. 4. The proportions of different classes for female and male,respectively

on the smooth curves be plus by a $\mathcal{N}(0;1)$ gaussian noise, so that Gaussian process can deal well with smooth curves, since convariance in Eq.(3) for a smooth curve is sometimes invertible. In Fig.2, we plot the smooth and noisy velocity and acceleration curves, where red curves stand for female, while green ones stand for male. As a result, our split EM algorithm clusters the velocity curves into 4 classes and the acceleration

Authorized licensed use limited to: Peking University. Downloaded on June 03,2022 at 10:03:45 UTC from IEEE Xplore. Restrictions apply.

curves into 3 classes, being shown in Fig.3, while Fig.4 shows the proportions of all the resulted classes for female and male, respectively.

In Fig.3a, the red mean function reaches a peak at age 14, which is later than the other three. Referring to Fig.4a, we can observe that the red class takes a considerable proportion in the male group than that in female group. Therefore, the time of pubertal spurt for most of the males is later than the females.

In Fig.3b, the blue class has a sharpest deceleration for velocity before age 4, and arrives a second sharpest deceleration for velocity at age 12 which is earlier than the other three. So, children in this class have a earlier time of pubertal spurt than the others. Referring to Fig.4b, we can observe that the blue class takes a larger proportion in the female group than in that male group. So, some cases of female's pubertal spurt arrive earlier than others, but rarely cases of male is such case.

Therefore, by the method whether the velocity curves are clustered as the 1st class to distinguish male or female, 47 out of 54 female are classified right, 36 out of 39 male are classified right. The misclassification error is 9.68%, which is smaller than 31.18% in [2].

VI. Conclusion

We have established a split EM algorithm for mixtures of Gaussian processes functional regressions which can make curve clustering effectively and efficiently. In fact, the Gauasian process functional regression model can learn the curve cluster more effectively and accurately. Moreover, the split mechanism based on the reversible jump theory under the MCMC framework and BIC makes the split EM algorithm effective with automated model selection. The experimental result on the both synthetic and real datasets demonstrate that our proposed split EM is robust and effective for curve clustering and even outperform the conventional EM algorithms.

Acknowledgment

This work was supported by the National Science Foundation of China under Grant 61171138.

References

[1] L. M. Sangalli, P. Secchi, S. Vantini, and V. Vitelli, "k-mean alignment for curve clustering," *Computational Statistics & Data Analysis*, vol. 54, no. 5, pp. 1219–1233, 2010.

[2] X. Liu and M. C. Yang, "Simultaneous curve registration and clustering for functional data," *Computational Statistics & Data Analysis*, vol. 53, no. 4, pp. 1361–1376, 2009.

[3] J. Q. Shi and B. Wang, "Curve prediction and clustering with mixtures of gaussian process functional regression models," *Statistics & Computing*, vol. 18, no. 3, pp. 267–283, 2008.

[4] J. Q. Shi, R. Murray-Smith, and D. M. Titterington, "Hierarchical gaussian process mixtures for regression," *Statistics & Computing*, vol. 15, no. 1, pp. 31–41, 2005.

[5] J. Q. Shi, R. Murray-Smith, and D. Titterington, "Bayesian regression and classification using mixtures of gaussian processes," *International Journal of Adaptive Control and Signal Processing*, vol. 17, no. 2, pp. 149–161, 2003.

[6] J. Shi, B. Wang, R. Murray-Smith, and D. Titterington, "Gaussian process functional regression modeling for batch data," *Biometrics*, vol. 63, no. 3, pp. 714–723, 2007.

[7] D. Wu and J. Ma, "A daem algorithm for mixtures of gaussian process functional regressions," in *International Conference on Intelligent Computing*. Springer, 2016, pp. 294–303.

[8] Z. Chen and J. Ma, "The hard-cut em algorithm for mixture of sparse gaussian processes," in *International Conference on Intelligent Computing*. Springer, 2015, pp. 13–24.

[9] D. Wu, Z. Chen, and J. Ma, "An mcmc based em algorithm for mixtures of gaussian processes," in *International Symposium on Neural Networks*. Springer, 2015, pp. 327–334.

[10] L. Zhao, Z. Chen, and J. Ma, "An effective model selection criterion for mixtures of gaussian processes," in *International Symposium on Neural Networks*. Springer, 2015, pp. 345–354.

[11] Z. Chen, J. Ma, and Y. Zhou, "A precise hard-cut em algorithm for mixtures of gaussian processes," in *International Conference on Intelligent Computing*. Springer, 2014, pp. 68–75.

[12] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.

[13] A. P. Benavent, F. E. Ruiz, and J. S. Martinez, "Ebem: An entropy-based em algorithm for gaussian mixture models," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 2. IEEE, 2006, pp. 451–455.

[14] A. Peñalver, F. Escolano, and B. Bonev, "Entropy-based variational scheme for fast bayes learning of gaussian mixtures," in *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. Springer, 2010, pp. 100–108.

[15] S. Richardson and P. J. Green, "On bayesian analysis of mixtures with an unknown number of components (with discussion)," *Journal of the Royal Statistical Society: series B (statistical methodology)*, vol. 59, no. 4, pp. 731–792, 1997.

[16] W. Zheng, Z. Ren, Y. Zhou, and J. Ma, "Byy harmony learning of log-normal mixtures with automated model selection," *Neurocomputing*, vol. 151, pp. 1015–1026, 2015.

[17] Z. Qiang and J. Ma, "Automatic model selection of the mixtures of gaussian processes for regression," in *International Symposium on Neural Networks*. Springer, 2015, pp. 335–344.

[18] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

[19] J. Ramsay, G. Hooker, and S. Graves, "Functional data analysis with r and matlab," *Use R*, vol. 8, no. 4, pp. 401–403, 2009.

Authorized licensed use limited to: Peking University. Downloaded on June 03,2022 at 10:03:45 UTC from IEEE Xplore. Restrictions apply.