Discussion: Clustering Random Curves Under Spatial Dependence

Gareth M. James, Wenguang Sun and Xinghao Qiao

Abstract

We discuss the advantages and disadvantages of a functional approach to clustering of spatial-temporal data. This leads us to suggest an alternative methodology which allows cluster memberships to vary over both temporal and spacial domains. One advantage of our approach is that it can easily incorporate time-varying covariates. A fitting algorithm is developed and we provide a simple simulation example to illustrate the performance of our method.

1 Introduction

We would like to congratulate the authors on an interesting and stimulating paper on spatially correlated clustering of functional data. Reading their work led us to consider the tradeoffs of the functional paradigm in relation to other strategies. In this discussion we summarize the functional clustering method and suggest an alternative approach to the problem.

The situation that is under consideration involves observed data, $Y_{ij} = Y(s_j, t_i)$, which is a realization of a spatial-temporal process where $S = \{s_1, \cdots, s_n\}$ is the collection of spatial locations and $T = \{t_1, \ldots, t_m\}$ is the collection of time points. Clustering could be performed directly on the $n \times m$ observations. Instead the authors have elected to model the data as functional in the time dimension which takes the view that the Y_{ij}'s are measurements of n spatially interdependent curves $\{Y_j(t), j = 1, \ldots, n\}$. Under this paradigm the data consists of n functional observations, each with a unique cluster membership; thus a cluster is defined as a collection of locations with similar temporal patterns. The goal is to divide the spatial domain into clusters by extracting information from the shapes of the underlying curves.

Clustering of functional data has been previously explored in a number of articles. However, most previous approaches have assumed that cluster memberships for each function are independent. The key contribution that the authors make is to incorporate a spatial correlation structure among the cluster memberships, $\mathbf{Z}_j = \{Z_{jc} : c \in C\}$, which are assumed to follow a multinomial distribution, where C is an index set and $Z_{jc} = I(\text{location } s_j \text{ is from cluster } c)$. In particular the authors propose the following Gibbs distribution to model $\mathbf{Z} = \{\mathbf{Z}_j : j = 1, \ldots, n\}$ as coming from a Markov random field (MRF),

$$P\left(Z_{jc} = 1 | \mathbf{Z}_{\partial j}\right) = \frac{1}{N_j(\theta)} \exp\left(\theta \sum_{i \in \partial_j} Z_{ij}\right), \tag{1}$$

where ∂_j denotes the spatial neighbors of s_j. This approach has the potential to improve the clustering accuracy by exploiting the local dependency among the cluster memberships.

In Section 2 we discuss some possible limitations of the functional approach and consider an alternative method where cluster membership can vary over both temporal and spatial domains. Our method has the advantage of allowing time-varying covariates to be included in the clustering process. Section 3 provides details on a fitting algorithm and some simple simulation results to illustrate the approach. We conclude in Section 4 with a brief discussion.

2 Temporal covariates and a spatial-temporal MRF

2.1 Some Limitations of the Functional Approach

One limitation of model (1) is that it does not allow for a spatial covariate, x_j, to be incorporated into the clustering approach. Conceptually, one could extend (1) to include x_j using the following model,

$$P\left(Z_{jc} = 1 | \mathbf{Z}_{\partial_j}, x_j\right) = \frac{1}{N_j(\theta_1, \theta_{2c})} \exp\left(\theta_1 \sum_{l \in \partial_j} Z_{lc} + \theta_{2c} x_j\right), \tag{2}$$

where θ_{2c} is a coefficient describing the effect of x_j on cluster c. Using model (2), a value of θ_{2c} that is large relative to the other clusters suggests that increasing values of x_j are associated with a higher probability of the jth curve belonging to cluster c. Similarly, the ratio of θ_1 and θ_{2c} controls the relative contributions of the spatial neighborhood versus the covariate in determining the cluster membership.

A more significant limitation of both models (1) and (2) is that they are unable to quantify the effects of time-varying covariates, or interventions on clustering. Time-varying covariates can be of great interest in some important applications. For example, a financial company may be interested in investigating whether a sequence of marketing strategies over time have significant effects on improving the profits of local stores in an area. Alternatively, in the context of the author's data, the local government may wish to know whether a new initiative over a given time period has effectively resulted in desirable temporal trends such as improved service quality in a community. Time-varying covariates cannot enter the MRF model because the cluster membership of location s remains fixed at all time points. Next we shall consider an alternative formulation of the clustering problem which can be used to address this limitation.

2.2 An Alternative Clustering Method

In some applications it may be desirable to characterize the cluster membership by jointly considering several spatial-temporal factors. For instance, in the context of the current paper we may wish to form a single clustering based on the accessibility to financial, medical, and educational services. Hence, we propose to extend the method in two directions, to model a collection of factors $\{1, \cdots, K\}$ and to allow the cluster memberships to vary in relation to both s_j and t_i, instead of only s_j as assumed in the functional approach.

In our formulation we wish to assign the cluster memberships according to a vector $\mathbf{Y}_{ij} = \{Y_{ijk} : k = 1, \cdots, K\}$ which is observed at location s_j and time t_i. Let $\mathbf{Z} = \{\mathbf{Z}_{ij} : i = 1, \ldots, m; j = 1, \ldots, n\}$ denote the cluster memberships where $\mathbf{Z}_{ij} = \{Z_{ijc} : c = 1, \cdots, C\}$ is assumed to follow a multinomial distribution with $Z_{ijc} = I(\text{point } (s_j, t_i) \text{ is from cluster } c)$. Then under this formulation, the conditional distribution of observations $\mathbf{Y}_{ij}$ given $\mathbf{Z}_{ij}$ is

$$\mathbf{Y}_{ij} | (Z_{ijc} = 1) = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_c + \boldsymbol{\epsilon}_{ijc}, \tag{3}$$

where $\boldsymbol{\mu}_0$ is the global effect, $\boldsymbol{\mu}_c$ is the cluster effect with constraint $\sum_{c=1}^{C} \boldsymbol{\mu}_c = 0$ and $\{\boldsymbol{\epsilon}_{ijc} : i = 1, \cdots, m; j = 1, \cdots, n\}$ are independent identically distributed (i.i.d.) multivariate normal variables with mean zero and variance-covariance matrix Σ_c.

Let ∂_{ij} denote the prescribed neighbors of point (s_j, t_i) in the spatial-temporal domain. At the initial time point, t_1, we have $\partial_{1j} = \{(s_l, t_1) : l \in \partial_j\}$, where ∂_j is the previously prescribed neighbors of location s_j in the spatial domain. At later time points, t_i, we have $\partial_{ij} = \{(s_l, t_i) : l \in \partial_j\} \cup \{(s_j, t_{i-1})\}$; hence the state of s_j will be influenced by both its spatial neighbors at time t_i as well as its previous state at time t_{i-1}. Generalizing the MRF approach we model the cluster memberships using the following spatial-temporal Markov random field (stMRF),

$$P\left(Z_{ijc} = 1 | \mathbf{Z}_{\partial_{ij}}\right) = \frac{1}{N_{ij}(\theta_1, \theta_2)} \exp\left(\theta_1 \sum_{l \in \partial j} Z_{ilc} + \theta_2 Z_{(i-1)jc}\right), \tag{4}$$

where $Z_{0jc} = 0$ for all j and c. Compared to the MRF defined by (1), the stMRF allows the cluster memberships to vary over both space and time but encourages $\mathbf{Z}_{ij}$'s that are adjacent in the spatial-temporal domain to have similar values. The parameters θ_1 and θ_2 control the level of spatial and temporal correlations, respectively. Large values of θ_2 will generate more stability over time with only minor changes in cluster membership while smaller values will allow faster updating of cluster memberships.

A significant advantage of the proposed stMRF is that it can be easily extended to handle time-varying covariates. For example, let x_{ij} denote a covariate of interest at location s_j and time t_i, then model (4) can be extended to,

$$P\left(Z_{ijc} = 1 | \mathbf{Z}_{\partial_{ij}}, x_{ij}\right) = \frac{1}{N_{ij}(\theta_1, \theta_2, \theta_{3c})} \exp\left(\theta_1 \sum_{l \in \partial j} Z_{ilc} + \theta_2 Z_{(i-1)jc} + \theta_{3c} x_{ij}\right). \tag{5}$$

Suppose that c^* is the *best* state and $x_{ij} > 0$ is an intervention at (s_j, t_i). Then, in a similar fashion to model (2), a value of θ_{3c^*} that is large relative to the other clusters suggests an effective intervention with increasing values of x_{ij} associated with higher probabilities of s_j belonging to cluster c^* at time t_i. Over time, locations with positive x_{ij} will drift towards cluster c^* with the rate of change determined by the size of θ_{3c^*}. In applications where we expect that it takes some time to observe the effect of an intervention, we may replace x_{ij} by a lagged value such as $x_{(i-1)j}$.

The clustering results based on the stMRF can be summarized and interpreted in different ways. We may recover interesting temporal patterns by investigating the underlying dynamic process $\mathbf{Z}_j = \{\mathbf{Z}_{ij} : i = 1, \ldots, m\}$ at location s_j. We can also identify interesting spatial patterns by investigating the process $\mathbf{Z}_i = \{\mathbf{Z}_{ij} : j = 1, \ldots, n\}$ at time point t_i. A 3-dimensional graphical representation of the cluster membership could be particularly useful to reveal the overall spatial-temporal trend.

3 Implementation

In this section we suggest an algorithm for fitting our model and illustrate the approach on a simple simulated data set.

3.1 Computational Algorithm

In contrast with the model considered by the authors which allows spatial dependence among $\mathbf{Y}_{ij}|\mathbf{Z}_{ij}$, we have limited our development to the case which assumes conditional independence of $\mathbf{Y}_{ij}|\mathbf{Z}_{ij}$. Denote by $\Theta_Y = (\boldsymbol{\mu}_0, \boldsymbol{\mu}_1, \cdots, \boldsymbol{\mu}_C, \Sigma_1, \cdots, \Sigma_C)$ the collection of observation distribution parameters and $\Theta_Z = (\theta_1, \theta_2, \theta_{31}, \cdots, \theta_{3C})$ the collection of Gibbs distribution parameters. Let $\mathbf{z}^*$ be the current estimates of $\mathbf{Z}$ and $\mathbf{y}$ be the observed values of $\mathbf{Y}$. Then we propose the following algorithm which iteratively updates the estimates of Θ_Y, Θ_Z and $\mathbf{Z}$ for the spatial-temporal model defined by (3) and (5).

1. Obtain an initial estimate for $\mathbf{Z}$ using a standard clustering approach which assumes independent $\mathbf{Y}_{ij}$.

2. Obtain $\hat{\Theta}_Y$ by maximizing the conditional likelihood

$$f(\mathbf{y};\Theta_Y|\mathbf{z}^*) = \prod_{i=1}^{n}\prod_{j=1}^{m} f(\mathbf{y}_{ij};\Theta_Y|\mathbf{z}^*_{ij}). \tag{6}$$

3. Obtain $\hat{\Theta}_Z$ which maximizes the following log pseudo-likelihood:

$$\log PL(\mathbf{z}^*;\Theta_Z) = \sum_{i=1}^{n}\sum_{j=1}^{m} \log P(\mathbf{Z}_{ij} = \mathbf{z}^*_{ij}|\mathbf{z}^*_{\partial_{ij}}, x_{ij}), \tag{7}$$

   where $P(\mathbf{Z}_{ij} = \mathbf{z}^*_{ij}|\mathbf{z}^*_{\partial_{ij}}, x_{ij})$ is defined in (5).

4. Update the estimate of $\mathbf{Z}$ based on current $\hat{\Theta}_Z$ and $\hat{\Theta}_Y$ using the iterative conditional modes (ICM) algorithm (Besag, 1986). Specifically, we obtain $\hat{\mathbf{z}}_{ij}$ by maximizing

$$P\left(\mathbf{Z}_{ij} = \mathbf{z}_{ij}|\mathbf{y}, \mathbf{z}^* \setminus \{\mathbf{z}^*_{ij}\}, x_{ij}\right) \propto f(\mathbf{y}_{ij}|\mathbf{z}_{ij})P(\mathbf{Z}_{ij} = \mathbf{z}^*_{ij}|\mathbf{z}^*_{\partial_{ij}}, x_{ij}) \tag{8}$$

   for each spatial-temporal point to complete one iteration of the ICM algorithm.

5. Go to step 2 until convergence.

Each step in this algorithm can be implemented using standard methods. In particular, the maximization of (6) in step 2 is simple because we can use the standard Gaussian MLE estimates for $\boldsymbol{\mu}_c$ and Σ_c, computed from the observations currently assigned to cluster c. In step 3 the parameters in the pseudo-likelihood function (7) can be computed using a baseline-category Logit model (see, for example, pp. 267 in Agresti, 2002). Finally, both right hand terms in step 4. have closed form expressions so (8) can be quickly maximized.

3.2 Simulation Results

To illustrate our approach we simulated data from the stMRF model on a 50 by 50 spatial grid at 10 time points with $\theta_1 = \theta_2 = 1$. We assumed three clusters; red, orange and white. The $\mathbf{Z}$'s were produced by first randomly generating cluster memberships at time point one, then iteratively resampling $\mathbf{Z}_{1j}$ at each location according to (5), conditional on the other cluster memberships. We repeated this for a reasonable burn in period until $\mathbf{Z}$ reached an equilibrium. The cluster memberships for the second and remaining time periods were then generated using a similar approach except that the cluster assignment from the previous period was also used. A time varying covariate, x_{ij}, was also included. To test out the covariate's effect over time we chose $x_{ij} = 0$ at time points 1 through 4 and then set $x_{ij} = 1$, over a subset of the spatial locations, for the remaining time periods. We used $\theta_{3\text{red}} = 1$ and $\theta_{3\text{orange}} = \theta_{3\text{white}} = 0$ so the covariate had the effect of increasing the probability of membership in the red cluster. Once all the $\mathbf{Z}_{ij}$'s were computed one dimensional $\mathbf{Y}_{ij}$'s were generated according to model (3).

The resulting cluster memberships for time periods 1, 5 and 10 are shown in the top row of Figure 1. The black squares in periods 5 and 10 illustrate the location where $x_{ij} = 1$ and demonstrate the tendency for these regions to end up in the red cluster. The plots illustrate a fairly high level of spatial correlation, especially in the later time periods. The stMRF model also demonstrates a clear evolution of cluster assignments over time. We use a standard package, "`mritc`" (Feng and Tierney, 2011) in R to fit a close approximation to our model. The main differences are that the package assumes $\theta_1 = \theta_2$ and it does not allow for the inclusion of extra covariates. The predicted clusters for periods 1, 5 and 10 using `mritc` are

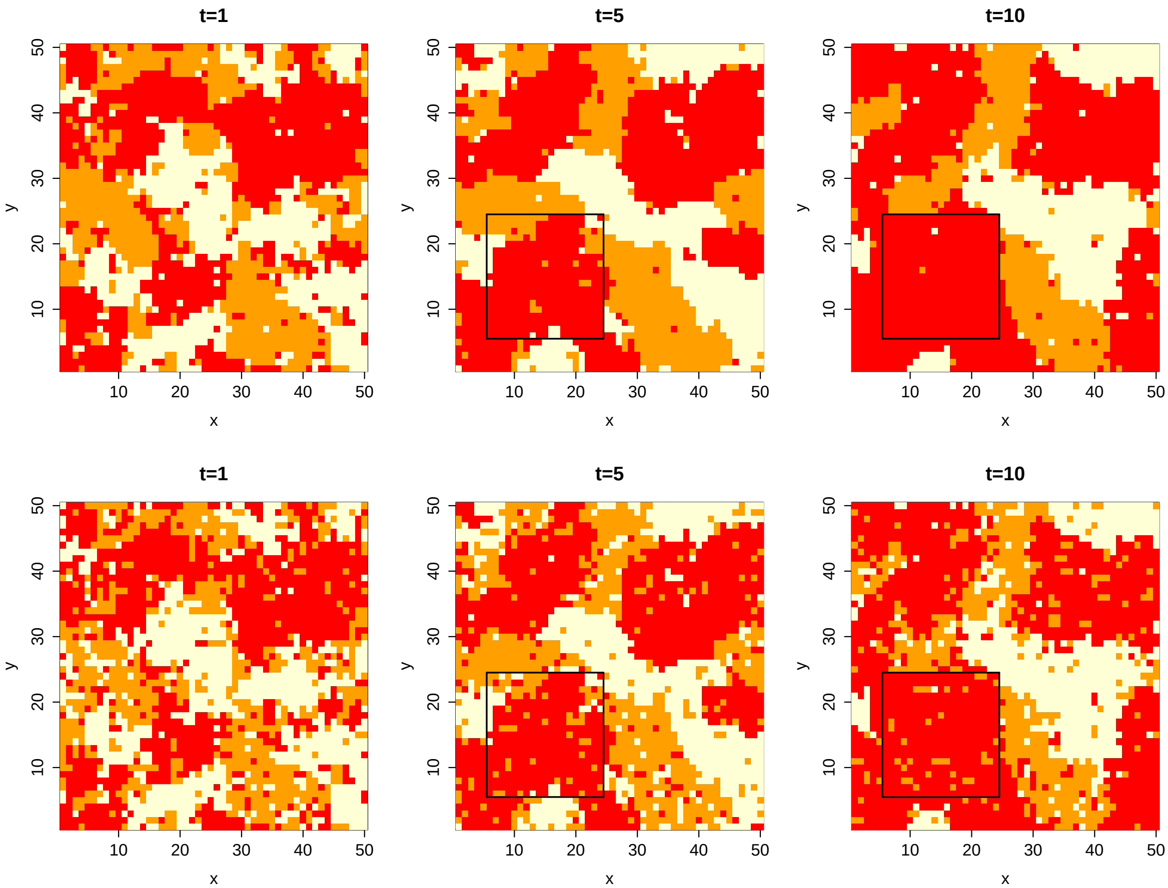

Figure 1: *Top Row: Simulated cluster assignments from the stMRF model at three of ten time points. Bottom Row: Estimate cluster assignments from the* `mritc` *function at the same time points.*

shown in the bottom row of Figure 1. Despite the fact that the model we are fitting does not incorporate the covariate information its effect clearly shows through and the method is still able to accurately identify the true clusters with an error rate of only 0.131.

4 Summary

The functional clustering approach views the realization of a spatial-temporal process as n spatially interdependent curves with cluster membership primarily characterized by the shapes of the curves. We propose an alternative formulation which models the spatial-temporal process using a spatial-temporal Markov random field model which allows the cluster memberships to vary in both spatial and temporal dimensions.

What are the tradeoffs between the two methods? The functional approach implicitly assumes that the process Y_{ij} varies smoothly over time and that the resulting curves can be grouped into a small number of similar clusters. For data of this type one might expect the functional approach to model the data more accurately than the stMRF method. Alternatively, our approach potentially provides some added flexibility in situations where there may be many different patterns of evolution in the temporal domain. The stMRF can potentially still model such data using a small number of clusters because cluster membership is allowed to vary over time. In addition the new approach makes it easier to examine the impact of time varying covariates on cluster assignment. On any given problem there are likely to be benefits from examining the data using both the functional and non-functional formulations.

References

Agresti, A. (2002). *Categorical data analysis* (Second ed.). Wiley Series in Probability and Statistics. New York: John Wiley & Sons.

Besag, J. (1986). On the statistical analysis of dirty pictures. *J. Roy. Statist. Soc. Ser. B 48*(3), 259–302.

Feng, D. and L. Tierney (2011). mritc: A package for mri tissue classification. *Journal of Statistical Software 44*, 1–20.