Spatio-temporal clustering of traffic networks

Ashwini Venkatasubramaniam[123], Ludger Evers[2],
Konstantinos Ampountolas[13]

[1] Urban Big Data Centre, University of Glasgow, United Kingdom,
[2] School of Mathematics and Statistics, University of Glasgow, United Kingdom,
[3] School of Engineering, University of Glasgow, United Kingdom

E-mail for correspondence: a.venkatasubramaniam.1@research.gla.ac.uk

Abstract: We present a novel Bayesian clustering method for spatio-temporal data observed on a network and apply this model to cluster an urban traffic network. This method employs a distance dependent Chinese restaurant process (DDCRP) to provide a cluster structure, by incorporating the observed data and geographic constraints of the network. However, in order to fully capture the dependency structure of the data, a conditional auto-regressive model (CAR) is used to model the spatial dependency within each cluster.

Keywords: network; spatial; clustering; Bayesian

1 Introduction

Heterogeneous urban traffic networks with regions of varying congestion levels have unique fundamental properties and *clustering* aids in the division of a city into homogeneous regions. We propose a novel Bayesian clustering technique for spatio-temporal network data which is based on an amalgamation of a distance dependent Chinese restaurant process (DDCRP) and a spatio-temporal conditional autoregressive model (CAR). We assume that we observe a time series of measurements that represent congestion levels aggregated over each junction in the network and the degree of similarity between adjacent junctions can be used to define spatially contiguous clusters. Existing literature relevant to clustering techniques for transportation networks account for spatial constraints but typically do not incorporate changes over time within a cluster. Traditional clustering algorithms such as k-means and probability mixture models also require choices to be made about the number of clusters.

---

2 Method

In our model, the road network forms an undirected graph with junctions acting as nodes and road segments between junctions as edges. This graph can be used to define the spatial component of the precision matrix. The spatial precision $\mathbf{\Sigma}_{\mathbf{S}}^{-1}$ within a cluster is modelled as a CAR model and the temporal precision $\mathbf{\Sigma}_{\mathbf{T}}^{-1}$ as an auto-regressive (AR1) model. Other models that incorporate temporal dependency such as the Matern covariance function are also possible. Due to the presence of a grid network topology with a limited number of road segments between junctions (typically not more than four per junction), the precision matrices exhibit sparsity and we utilize a CAR model proposed by Leroux (2000) to define this spatial precision. We define an adjacency matrix $\mathbf{W}$ and the precision matrix $\mathbf{Q} = \mathbf{\Sigma}_{\mathbf{S}}^{-1} = \rho(\text{diag}(W_{k++}) - \mathbf{W}) + (1-\rho)\mathbf{I}_{n_J}$, where ρ controls how strongly correlated adjacent junctions are, $\text{diag}(W_{k++})$ is a diagonal matrix with elements equivalent to the row sums of $\mathbf{W}$, and $\mathbf{I}_{n_J}$ is an $n_J \times n_J$ identity matrix ($n_J =$ number of junctions). With the presence of a unique observation for every space-time combination (n_J junctions and n_T time points), a covariance matrix $\mathbf{\Sigma}$ can be written as $\mathbf{\Sigma}_{\mathbf{S}} \otimes \mathbf{\Sigma}_{\mathbf{T}}$. Clusters can be obtained by removing edges such that the graph can be partitioned into components not connected to each other. In Figure 1, a network composed of eight nodes can be divided into two clusters such that there are road segments but no links between adjacent junctions of two differing clusters.

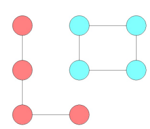

FIGURE 1. Graph showing two clusters formed in the network based on the presence or absence of a road segment between junctions.

Individually removing edges not supported by the data would yield a sparse graph, but would be unlikely to result in a graph with more than one component. Instead, we use a prior to enforce that edges are omitted in a way that leads to a clear partitioning of the graph. We utilize a modified version of the DDCRP first introduced by Blei (2011) that allows our model to incorporate geographic constraints of the network, account for the shape, and determine the number of clusters. The DDCRP makes assumptions of non-exchangeability to account for components of distance such as time, space, etc. In a traditional Chinese restaurant process (CRP) (also a special case of the DDCRP), a restaurant can be assumed to consist of an infinite number of tables. Customers $i = 1 \ldots n$ are individual data points that enter and take a seat at a randomly chosen table k. Tables are deemed to be clusters and after a finite number of customers n_k have been seated, the seating plan represents a partition. In the usual representation of the

CRP, customers choose tables. In a DDCRP, clusters are instead determined based on friendships between customers i and j, with a group of friends then sitting at an assigned table z_i, i.e., forming a cluster. Thus, $z(\mathbf{c})$ are table assignments that follow from customer assignments. In a non-sequential DDCRP, clusters arise from some customers choosing to befriend themselves or someone already connected to them, resulting in a redundant assignment. We also modify the DDCRP to allow customers to befriend more than one customer, which controls for the number of singleton clusters. Let c_i be the index of a customer that is sitting with customer i and we describe the distribution of this customer assignment as:

$$P(c_i = j|\alpha) \propto \begin{cases} 0, j \neq i \text{ and } i \nsim j \\ 1, j \neq i \text{ and } i \sim j \\ \alpha, j = i \end{cases}$$

In our method, customers are junctions and friendships can only occur along road segments between junctions. Since this modified DDCRP suggests a prior over a combinatorial number of junction assignments, the posterior is intractable and inference is carried out using a Metropolis within Gibbs sampler. We assume that the measure of congestion levels $\mathbf{Y}$ follows a Gaussian distribution and the likelihood at partition $z(\mathbf{c})$ gives the product of probabilities calculated for sets of observations at each determined cluster. To account for the spatial dependency within the cluster, we define $\mathbf{\Sigma_S}$ and $\mathbf{\Sigma_T}$ to represent the precision matrix described earlier. Accordingly, the likelihood can be defined as:

$$\begin{aligned} \ln(P(\mathbf{Y}|\mathbf{\Sigma_S}, \mathbf{\Sigma_T}, \sigma^2, \tau^2)) = &- \frac{n_J n_T}{2} \ln(2\pi) - 0.5 \ln|\sigma^2 \mathbf{I} + \tau^2 \mathbf{\Sigma_S} \otimes \mathbf{\Sigma_T}| \\ &- 0.5 \operatorname{vec}(\mathbf{Y})^{\mathrm{T}} [\sigma^2 \mathbf{I} + \tau^2 \mathbf{\Sigma_S} \otimes \mathbf{\Sigma_T}]^{-1} \operatorname{vec}(\mathbf{Y}) \end{aligned}$$

We can rewrite terms in the likelihood, $\operatorname{vec}(\mathbf{Y})^{\mathrm{T}}[\sigma^2\mathbf{I}+\tau^2\mathbf{\Sigma_S}\otimes\mathbf{\Sigma_T}]^{-1}\operatorname{vec}(\mathbf{Y}) = \operatorname{vec}(\mathbf{Y})^{\mathrm{T}}(\mathbf{\Gamma_S}\otimes\mathbf{\Gamma_T})(\sigma^2\mathbf{I}+\tau^2\mathbf{\Lambda})^{-1}(\mathbf{\Gamma_S^{\mathrm{T}}}\otimes\mathbf{\Gamma_T^{\mathrm{T}}})\operatorname{vec}(\mathbf{Y}) = \operatorname{vec}(\mathbf{\Gamma_T^{\mathrm{T}}}\mathbf{Y}\mathbf{\Gamma_S})^{\mathrm{T}}(\sigma^2\mathbf{I}+\tau^2\mathbf{\Lambda_S}\otimes\mathbf{\Lambda_T})^{-1}\operatorname{vec}(\mathbf{\Gamma_T^{\mathrm{T}}}\mathbf{Y}\mathbf{\Gamma_S})$, where σ^2 is variance of the noise, τ^2 is a prior variance, $\mathbf{\Lambda_T}$ represents a diagonal matrix of the eigenvalues of $\mathbf{\Sigma_T}$, and $\mathbf{\Gamma_T}$ represents a matrix of the eigenvectors of $\mathbf{\Sigma_T}$. Here, we only need to compute the diagonal of $[\sigma^2\mathbf{I} + \tau^2\mathbf{\Lambda_S} \otimes \mathbf{\Lambda_T}]^{-1}$ on the rotated data $\operatorname{vec}(\mathbf{\Gamma_T^{\mathrm{T}}}\mathbf{Y}\mathbf{\Gamma_S})^{\mathrm{T}}$. In addition, $\ln|\sigma^2\mathbf{I}+\tau^2\mathbf{\Sigma_S}\otimes\mathbf{\Sigma_T}|$ can be rewritten as $\sum_{t=1}^{n_T}\sum_{s=1}^{n_J} - \ln(\tau^2\mathbf{\Lambda_T}[t,t]\cdot\mathbf{\Lambda_S}[s,s]+\sigma^2\mathbf{I})$. Together, these terms can be evaluated for an efficient solution in $O(n_J^2 n_T + n_T^2 n_J + n_J^3 + n_T^3)$ rather than $O(n_J^3 n_T^3)$. Sampling from this posterior can happen in two phases where we first remove the customer and then consider how the likelihood term can be changed when this customer is replaced. The sampler thus has the potential to change multiple cluster assignments through a single change in customer assignment and using these moves is able to explore the space of possible partitions to determine a partition structure conditional on observed data.

3 Results

Occupancy is defined as the percentage of time that a location on the road is occupied by vehicles. In our example, occupancy data was generated using the AIMSUN microscopic simulator for a network in downtown San Francisco composed of 316 links and 158 junctions. This was recorded over a period of six hours with a sampling frequency of 180 seconds. We cluster the simulated network such that each individual cluster represents a level of occupancy that is distinct from other clusters, as shown in Figure 2.

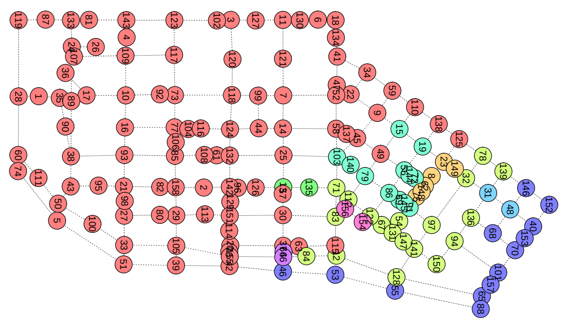

FIGURE 2. Traffic network with clusters that indicate different congestion levels.

This paper proposes a Bayesian clustering algorithm that accounts for spatial constraints and is modelled in a computationally efficient manner on data with varying temporal patterns. Further work seeks to identify clusters that change over time. However, current Kronecker product tricks that enhance efficiency cannot be utilized since spatial precision would change over time.

Acknowledgments: Research funded by the Lord Kelvin Adam Smith scholarship, University of Glasgow, 2014–2018

References

Blei, D. and Frazier, P. Distance dependent Chinese restaurant processes. *Journal of Machine Learning Research*, 12(Aug):2461–2488, 2011.

Leroux, B., Lei, X., and Breslow, N. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In *Statistical models in epidemiology, the environment, and clinical trials*, pages 179–191. Springer, 2000.