Supplementary Material

Gaussian Process Subset Scanning for Anomalous Pattern Detection in Non-iid Data

William Herlands Carnegie Mellon University
Edward McFowland III University of Minnesota
Andrew G. Wilson Cornell University
Daniel B. Neill Carnegie Mellon University

1 Alternative model MLE

Given data, (x, y), we can determine the optimal mean shift, β^* through maximum likelihood estimation as shown below. Let μ, Σ be the posterior mean and covariance of the null model in the domain of x, and denote $E = \Sigma^{-1}$ for brevity.

$$\begin{aligned}\beta^* &= \max_{\beta}\Big((2\pi)^{-\frac{k}{2}}|\Sigma|^{-\frac{1}{2}}\exp(-\frac{1}{2}(y - w\beta - \mu)^T \\ &\qquad E(y - w\beta - \mu))\Big) \\ &= \max_{\beta} -\frac{1}{2}(y - w\beta - \mu)^T E(y - w\beta - \mu) \\ &= \max_{\beta}(y-\mu)^T E w\beta - \frac{1}{2}(w\beta)^T E(w\beta)\end{aligned} \tag{1}$$

We take the derivative with respect to β and set it to zero

$$\begin{aligned}\frac{\delta LLR(w)}{\delta\beta} &= (y-\mu)^T E w - (w\beta^*)^T E(w) = 0 \\ &\Rightarrow (w\beta^*)^T E(w) = (y-\mu)^T E w \\ &\Rightarrow \beta^* = \frac{w^T E(y-\mu)}{w^T E w}\end{aligned} \tag{2}$$

2 Iterative β_{MAX} algorithm to approximate optimal subset

Since the derivation of β_{MAX_i} is conditional on a subset w, we obtain the *conditional* optimal subset. In order to approximate an optimal solution we use iteratively compute the conditional optimal subset beginning with a null subset, $w = \vec{0}$. This is an $O(\ell k \log(k))$ algorithm for some ℓ number of iterations, where k is the size of the neighborhood. Pseudo-code is depicted in Alg. 1.

Algorithm 1: Iterative β_{MAX_i} algorithm

```
Result: Highest scoring subset $w^*$
Initialize $w = \vec{0}$;
for $l = 1 : \ell$ do
    Compute $\beta_{MAX_i}$ $\forall i$ conditioned on the current value of $w$;
    Find highest scoring subset, $w^{(l)}$, using a linear search over sorted $\beta_{MAX_i}$;
    Compute $LLR(w^{(l)})$;
    Set $w = w^{(l)}$;
end
Choose $w^* = \arg\max_{w^{(l)}} LLR(w^{(l)})$
```

3 Constrained β_{MAX} optimization over blocks

Although we focus on unconstrained subsets searching within neighborhoods, real world applications sometimes require a more constrained optimization. For example, in spatiotemporal phenomena it is often useful to consider anomalous patterns that are nearby in space and contiguous over time. We can enforce such constraints by predefining mutually exclusive blocks of points, $(x^{(B)}, y^{(B)}) \subseteq (x^{(n)}, y^{(n)})$ where points in a block must all either be included in, or excluded from, a subset.

When considering blocks of points we can compute the total contribution from all points in the block, though we must also account for additional off-diagonal terms in E due to the blocking of data points. Following the derivation in Section 4.2.1 of the main paper, we can derive the β_{MAXb} for each block,

$$\beta_{MAX_B} = \sum_{i \in B} \frac{2\big(E(y^{(n)} - \mu)\big)_i}{\big(\sum_{j \notin B} 2 w_j E_{j,i} + E_{i,i} + \sum_{k \in B} E_{k,i}\big)} \tag{3}$$

This can be used in a lightly modified version of Algorithm 1 where the β_{MAXb} of blocks, not individual points, is iteratively computed.

4 School Absenteeism

Public schools in New York City record and publish daily student attendance (NYC Department of Education, 2017). Given the importance of education on future outcomes there is tremendous interest in understanding patterns of school absenteeism. We consider public school attendance data in Manhattan for the 2015-2016 school year. The data is messy, with missing entries and non-uniform placement of school locations. We aggregate data at weekly level and remove the last four weeks of the school year since they contain known high absenteeism rates that are not of interest to Department of Education officials.

We apply GPSS methods and baseline approaches with neighborhoods of up to ten local schools. All GPSS methods identified an anomaly around January to February 2016 concentrated on West Side of Manhattan. The results from GRQ around the time of the detected anomaly are presented in Fig. 1. Each dot represents a school location, with yellow dots indicating high attendance and blue dots indicating low attendance. The space-time locations of schools in the top ten anomalous subsets are bordered in red.

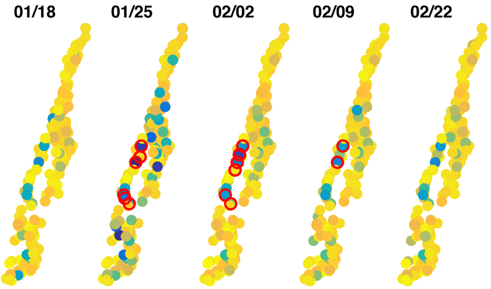

Figure 1: School absenteeism results from Manhattan using GRQ. Each dot represents a school location, with yellow dots indicating high attendance and blue dots indicating low attendance. The space-time locations of schools in the top ten anomalous subsets are bordered in red.

The detected anomalies correspond to a category five blizzard which may have disrupted teachers and students from attending school even though no snow day closings were reported at the time. Further research is required to understand why the West Side of Manhattan differed systematically from the rest of the borough. Baseline anomaly detection methods did not identify a coherent anomaly and instead detected anomalies throughout the year.

References

NYC Department of Education. Data about schools, 2017. URL http://schools.nyc.gov/AboutUs/schools/data/Attendance.htm.