RaspBary: Hawkes Point Process Wasserstein Barycenters as a Service

Ryan Hosler and **George Mohler**
Indiana University - Purdue University Indianapolis
rjhosler@iu.edu, gmohler@iupui.edu

March Saper
Olin College of Engineering
march.saper@students.olin.edu

Abstract

We introduce an API for forecasting the intensity of space-time events in urban environments and spatially allocating vehicles during times of peak demand to minimize response time. Our service is applicable to dynamic resource allocation problems that arise in ride sharing, mobile delivery, emergency vehicle placement, etc. We illustrate the service using medical emergency data from Indianapolis and show that the system can predict future medical emergencies and allocate ambulances such that response time is significantly decreased. The software is available for public use by app developers with a need for vehicle allocation over spatial-temporal data in real-time.

Introduction

Software applications that deal with urban logistics must solve the coupled problem of i) estimating a space-time intensity of events (demand) and ii) allocating vehicles to the dynamic intensity to match demand and minimize response time. Hawkes processes have recently become a popular model for solving i), however scalable Hawkes process software is lacking that can be easily integrated in mobile and web apps (see (Reinhart 2017) for a review of the space-time Hawkes process literature). We introduce a software application that allows for scalable training and prediction of spatial Hawkes processes using online gradient descent to solve i). Secondly, Wasserstein barycenters have recently been introduced for minimizing the earth mover's distance between probability densities (Cuturi and Doucet 2014). We incorporate fast Wasserstein barycenters computation into the API service for solving the optimal vehicle allocation problem. Both the Hawkes process and Wasserstein barycenters are implemented as a model service accessible via GET and POST requests. We illustrate the service using Indianapolis medical emergency data, showing that the average distance to demand can be reduced from 1.95 to 1.27 miles.

System and Methods

Figure 1 shows our system. Our modeling and clustering methods are accessible through modular API calls (code available at (Hosler and Saper 2018)):

- A GET request sends query parameters and receives the desired intensity data. A start time parameter initiates when to begin predictions and an interval count parameter determines the amount of predictions to make (at specifiable intervals into the future).
- A POST request returns optimal vehicle placements. The request specifies current vehicle locations in latitude and longitude coordinates and whether that vehicle can be moved for optimization. An interval count and start time parameter determine what Hawkes intensities to use for clustering.
- A GET request for posting new events for online model training or CSV upload for batch model updates.

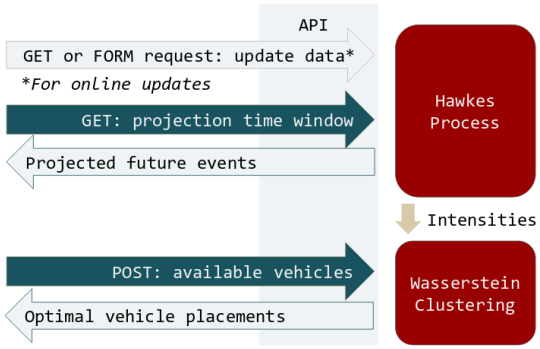

Figure 1: System diagram of our API service.

The Hawkes process method consists of non-parametric estimation of a background intensity capturing time of day and day of week trends in addition to spatial heterogeneity. We use a non-parametric model for the triggering kernel to capture contagion in events. In particular we use an exponential basis for the Hawkes kernel that facilitates fully online gradient descent training (Mohler and Brantingham 2018).

To allocate vehicles, we utilize a fast computational approach (Cuturi and Doucet 2014) to minimize Wasserstein distance to the Hawkes process intensity. Each vehicle is represented as a delta function density and the density location is optimized to minimize earth mover's distance to the Hawkes process. Wasserstein barycenters yield clusters with similar demand levels, in contrast to k-means where each vehicle location (centroid) may have a different demand volume to serve.

Demonstration

In this section we illustrate the capabilities of our service using data provided by Indianapolis Emergency Medical Services (EMS).

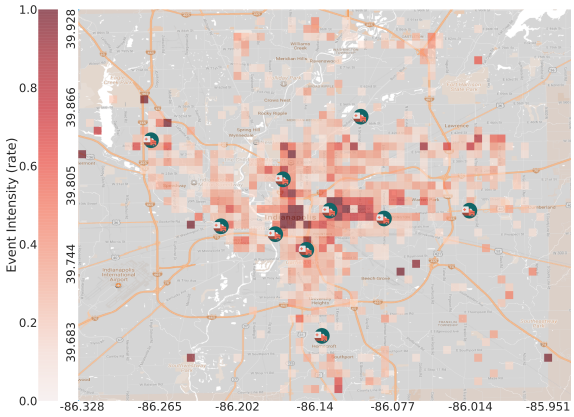

Figure 2: Example intensity projection and ambulance allocation in Indianapolis over a representative 4 hour period.

The Predictive Accuracy Index (PAI) (Mohler and Brantingham 2018) measures the percentage of incidents captured in the top k grid cells flagged as areas of peak activity. The PAI is area normalized so that a value of 1 corresponds to random predictions. 50 test projections each over periods of 1, 2, and 4 hours yielded an average predictive accuracy index of 17.1 when compared with corresponding historical data. These results are shown in Figure 3.

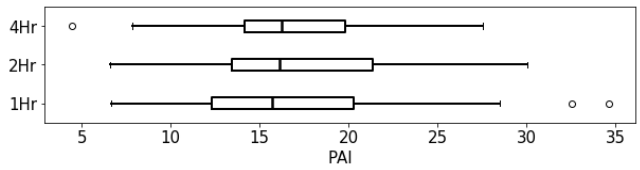

Figure 3: Predictive accuracy index for 50 simulations over 1, 2, and 4 hours.

Peak event time prediction is an important application in ambulance allocation as cities have flex trucks that can be dynamically positioned, but these trucks are parked at stations during low volume hours (for driver considerations). Figure 4 shows 2 weeks of projections and the corresponding historical data examining the total number of events that occurred in 2 hour periods. We defined peak event activity as a period during which the number of events fell in the 80th percentile or higher. We were able to predict these periods with an AUC of 0.721.

To test optimal ambulance placement, we ran 3 simulations allocating 50 ambulances: not moved from the EMS stations (2 ambulances per station), clustered by Kmeans, clustered by Wasserstein barycenters. Each simulation tested the response distance to 50 real emergencies that took place during the predicted period. Every 10 emergencies, the ambulances were reallocated. The average driving distances

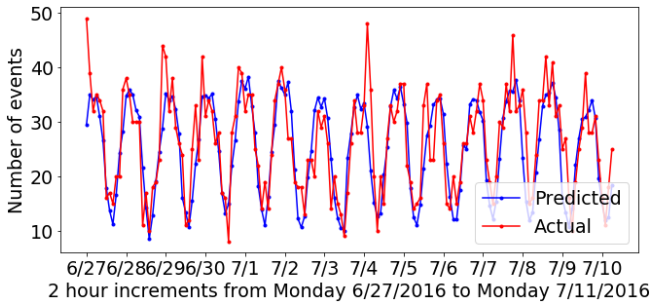

Figure 4: Total number of events in all grid cells in 2 hour windows over a 2 week period.

were 1.946 mi with no ambulance allocation, 1.400 mi with Kmeans and 1.268 mi with Wasserstein barycenters. The full results of this simulation are shown in Figure 5.

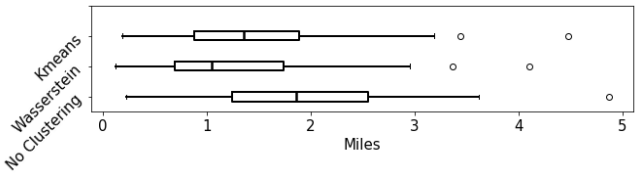

Figure 5: Driving distance to 50 real events that occurred in test projection period.

Our API is extensible to a wide variety of spacial-temporal data. The trend modeling and point process based clustering enable development of applications far superior to manually allocating resources on a heat map, the solution to which many government organizations still resort. Based on our work with Indianapolis EMS data, we have demonstrated the efficacy of our service and we believe it can extend beyond EMS applications in Indianapolis.

Acknowledgements

This work received support from NSF grant SCC-1737585, NSF grant ATD-1737996 and NSF grant REU-1343123.

References

Cuturi, M., and Doucet, A. 2014. Fast computation of wasserstein barycenters. In Xing, E. P., and Jebara, T., eds., *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, 685–693. Bejing, China: PMLR.

Hosler, R., and Saper, M. 2018. Hawkes process wasserstein barycenter github repository. `https://github.com/rjhosler/IUPUI-REU/`.

Mohler, G., and Brantingham, P. J. 2018. Privacy preserving, crowd sourced crime hawkes processes. In *2018 International Workshop on Social Sensing (SocialSens)*, 14–19.

Reinhart, A. 2017. A review of self-exciting spatio-temporal point processes and their applications. *Statistical Science*.