A PIPELINE FOR EMERGENCY RESPONSE

Ayan Mukhopadhyay
Electrical Engineering & Computer Science
Vanderbilt University
Nashville, TN, USA
ayan.mukhopadhyay@vanderbilt.edu

Yevgeniy Vorobeychik
Computer Science & Engineering
Washington University in St. Louis
St. Louis, MO, USA
yvorobeychik@wustl.edu

ABSTRACT

The problem of ensuring timely response to spatial-temporal incidents like crimes, accidents, fires and distress calls is of utmost importance to urban communities across the globe. This problem can actually be broken down into the following modular sub-problems - understanding where and when such incidents occur, how resources should be placed in anticipation of such incidents, and finally, how should such resources be dispatched as and when needed. While all these problems have attracted interest from various communities, there are significant prior gaps in the literature. Further, we argue that any holistic solution to this problem must look at creating a pipeline where solutions to these subproblems can work in harmony. In this paper, we look at these problems individually in a systematic and principled manner, and then evaluate their functioning as a pipeline. We also demonstrate how our approach outperforms state-of-the-art approaches using real-world data from the metropolitan area of Nashville, USA.

1 INTRODUCTION

In the last century, a rather alarming increase in population density across the globe has led to a number of challenges, out of which crimes and accidents are two issues that plague most of the urban regions of the world. To tackle such problems, government bodies resort to a variety of emergency responders including police, medical and fire departments. There are three problems that are vital to such responders - 1) understanding *where* and *when* such incidents happen, and 2) how can responders be placed to best respond to such incidents, and 3) how can such responders be deployed in real-time when such incidents occur. Our approach towards tackling the problem of emergency response consists of two major steps - first, we seek to develop rigorous and principled approaches to solve each sub-problem independently, and secondly, we aim to create a pipeline that can merge the solutions to create a holistic and complete framework for ensuring effective emergency response.

We begin by trying to understand the dynamics of such spatial-temporal incidents. Such forecasting models have been extensively studied; for example, both qualitative and quantitative approaches have been studied to forecast crimes as well as traffic accidents (Speer et al., 1998; Kennedy et al., 2011; Zhang et al., 2015; 2016; Balke et al., 2005; Ackaah & Salifu, 2011). A major shortcoming of such prior approaches is that they do not allow principled data-driven continuous-time incident forecasting that can include arbitrary risk factors. Mukhopadhyay et al. (2016) tackle these problems efficiently, by using parametric survival models (Cox & Oakes, 1984) to predict incidents like crimes. Parametric survival models are flexible to accommodate an arbitrary set of features, and can be learned using principled methods; moreover, such a model shows significant improvement over existing state-of-the-art. However, such an approach has a few shortcomings - first, it treats all incidents with equal severity or importance, which is detrimental to planning strategies of emergency response. Also, the spatial granularity learning such a model is based on ad-hoc heuristics, that might not work on all types of incidents or regions. In this paper, we explain how these gaps can be systematically bridged.

Given a model to forecast incidents, we then move to the problem of dispatching responders when incidents occur. Solving the problem of optimally dispatching responders is incumbent upon first placing the responders effectively in anticipation of such incidents. We assume the availability of such an approach, and briefly describe one such method in section 4. Given such an initial distribution of resources, the canonical way of solving the dispatching problem is to model it as a Semi-Markov Decision Problem (SMDP) (Keneally et al., 2016; Mukhopadhyay et al., 2018). This formulation is challenging to solve, since the state-transition probabilities of the SMDP are difficult to calculate in closed-form. Mukhopadhyay et al. (2018) present the more general of the two approaches, by first converting the SMDP to a discrete-time MDP, and accessing a simulator to learn the state-transition probabilities. Although such an approach is guaranteed to converge to the optimal solution, it can not work in dynamic settings, and moreover, the policy takes weeks to learn. In order to alleviate this problem, we design an efficient approach that is fast and scalable, and can compute close-to-optimal solutions even in dynamic environments. Finally, we use real-world data from Nashville, a major metropolitan area of USA, to compare our approach against existing state-of-the-art, and show that our models significantly outperform them.

2 Problem Setup

We consider a set of equally sized grids G, that span the entire spatial area under consideration. Also, we assume the availability of a dataset D of prior incidents, in which incident $d_i \in D$ denotes the i^{th} incident. Each incident is labeled with its location of occurrence, which is denoted by the gird in which it happened. Further, each incident $d_i \in D$ is also characterized by a time of occurrence τ_i, a reported severity that is represented by the discrete ordinal random variable k_i, and a set of spatial-temporal features $w_i \in \mathbb{R}^m$. Thus, our dataset is a collection of time-stamped feature vectors corresponding to crime incidents, $\{(\tau_1, k_1, w_1), (\tau_2, k_2, w_2), \ldots, (\tau_n, k_n, w_n)\}$. For each grid $g_i \in G$, we denote the time between successive incidents by the random variable t, and define $t_i = \tau_i - \tau_{i-1}$ as the time to arrival of the i^{th} incident in the dataset. Our primary goal is to learn a distribution $f(t, k|w)$.

Given such a model, we consider a set R of responders, that are spatially distributed in G. To solve the dispatch problem, we seek to learn a policy π that can prescribe an action (which responder to dispatch) when an incident occurs.

3 Incident Forecasting

3.1 Predicting Incident Arrival

We first present a model that can be used to efficiently learn a marginal distribution over inter-arrival times ($f_t(t|w)$), and then tackle the problem of learning the joint distribution. A natural fit for this problem is survival analysis, which has been used to predict urban crime incidents (Mukhopadhyay et al., 2016). Survival Analysis comprises of a broad class of methods that model the distribution of time to an incident arrival. Formally, a survival model is $f_t(t|\gamma(w))$, where f_t is a probability distribution for a continuous random variable T representing the inter-arrival time, which typically depends on covariates w as $\log(\gamma(w)) = \theta_0 + \sum_i \theta_i w_i$. In order to model and learn $f(t)$, we chose the exponential distribution, which has been widely used to model inter-arrival time and has recently been used to predict urban incidents. Given such a model and a dataset D, the goal of a learner is to find parameters θ^* that maximize the sum of log-likelihood of the inter-arrival times, such that $\theta^* = \text{argmax}_{\theta \in \Theta} \sum_{i=1}^{n} \log f_t(t_i|\gamma(w_i))$, which can be solved by standard gradient-based methods.

This basic machinery of survival modeling can be used to forecast spatial temporal incidents across an urban area. However, we now seek to understand the granularity at which such a model should be learned. One obvious way is to learn such a survival model independently for each grid $g_i \in G$. The main concern with this approach is overfitting: each grid induces relatively little data, and there are surely considerable structural similarities of the incident process across multiple grids that we can leverage. On the other hand, learning a single "universal" model for all grids may fail to capture all of the existing heterogeneity not explicitly modeled in the feature space w. We present a principled way of tackling this problem by using a Hierarchical Clustering approach (Johnson, 1967). Consider the feature set w to be divided into two parts, namely w_s and w_d, where w_s represents a set of *static* features, such as population density in a grid, which remain relatively fixed for an area, while w_d denotes *dynamic* features, such as the amount of rainfall in a day or day of the week. Our fundamental hypothesis is that the set of static features w_s can be used to identify similarities between distinct spatial grids. This hypothesis lets us create clusters of *similar* grids, and then one survival model can be learned per cluster. To operationalize this hypothesis, we propose a hierarchical clustering algorithm, that we describe at a high level for the sake of brevity. At the onset, we treat each grid as a distinct cluster. Iteratively, we merge two grids that are most similar, with similarity between grid i and grid j measured as the squared norm of the associated feature vectors w_s^i and w_s^j. At each step, we check whether the updated set of clusters decreases the predicted likelihood computed on the training data set compared with the previous iteration by more than a pre-defined limit, and stop as soon as this improvement in likelihood is below a pre-specified limit.

3.2 Predicting incident severity

One way of capturing incident severity is to learn a different arrival model for each severity type. However, this approach has shown that sacrificing the scale of the data for achieving heterogeneity can create noisy estimates (Gorr et al., 2003). In order to tackle this problem, we decompose the joint distribution $f(t, k|w)$, and represent it as $f(t, k|w) = f(t|w)f(k|t, w)$. This decomposition helps us in two ways: first, the parametric survival model described in section 3.1 can directly be used to learn the density of inter-arrival times, and secondly, the entire dataset can be used to learn a model over incident severity, without fracturing it by severity category. To learn the severity distribution conditional on incident time and the feature vector w ($f(k|t, w)$), we use the multinomial logistic regression model (Böhning, 1992).

4 Dynamic Dispatch of Responders

The efficacy of any approach to dynamically dispatch responders is heavily dependent upon first placing the available responders in anticipation of such incidents. One such way of allocating resources given a model of incident arrival is to minimize the expected waiting time of the future incidents. The resulting optimization problem is non-linear and non-convex, but can be solved well with well-crafted meta-heuristics (Mukhopadhyay et al., 2017). Given an allocation of responders and a model of incident arrival, the problem of dynamic dispatch of responders can be formulated as a semi-Markov decision process (SMDP), and then solved by converting the problem to a discrete-time MDP. This process, however, is extremely slow and fails to work in dynamic environments since any change in the problem definition (the number of responders, or the position of a depot) renders the learned policy stale.

In order to tackle this problem, we first highlight an important observation - one need not find an optimal action for each state as part of the solution approach since at any point in time, only one decision-making state might arise in a system that evolves in continuous time. This difference is crucial, as it lets us bypass the need to learn an optimal policy for the entire MDP. Instead, we focus on a principled approach that evaluates different actions at a given state, and selects one that is sufficiently close to the optimal action. We do this using sparse sampling, which creates a sub-MDP around the neighborhood of the given state and then searches that neighborhood for an action. In order to actualize this, we use Monte-Carlo Tree Search (MCTS). Once an incident happens, possible actions are considered based on the responders that are available. Then, a tree is gradually built by simulating each action and the entire area under consideration. We point out that the generative incident arrival model is particularly vital in this regard, since future incidents can be forecasted easily by such a model. Finally, the action that maximizes the long-term rewards for the system is chosen (in order model, shorter response times ensure higher rewards).

5 Experimental Findings

We test our algorithmic approach on real-world data from a major metropolitan area of USA, with a population of approximately 700,000. We looked at data for 26 months, from 2014 - 2016, comprising of a total of 20148 traffic accidents. We divide the data into 3 overlapping datasets for our experiments. For the incident prediction model, we used prior work (Songchitruksa & Balke, 2006; Poch & Mannering, 1996; Shankar et al., 1995) and expert opinions to construct the feature set w. We took into effect weather parameters (rain and snow), temporal cycles (time of day, day of week), temporal and spatial incident correlation (number of incidents in grids and neighboring grids) and transportation features (number of roadway and highway intersections in a grid).

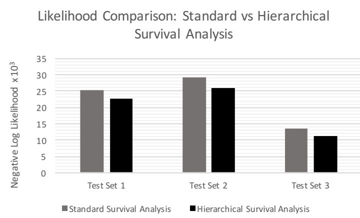

Figure 1: Likelihood Comparison (Lower is better)

Test Set	Percent Accuracy	False Negative	False Positive
Set 1	64.7%	3.4%	31.9%
Set 2	65.4%	1.6%	33%
Set 3	67.4%	1.8%	30.8%

Table 1: Severity Prediction Accuracy

Number of Stations (Fraction of Nashville Count)	26 (full)	13 (1/2)	6 (1/4)	3 (1/8)
Average Response Time Savings for Incidents Impacted by Policy (seconds)	38.705	2.231	15.917	34.871
Number of Incidents Impacted by Policy	5	14	99	150
Average Computation Time per Incident (seconds)	0.384	0.198	0.350	0.0343

Table 2: Performance of System Compared to Base Policy

To evaluate the efficacy of our incident prediction model, we compare Hierarchical Survival Analysis (HSA) with a baseline survival model, that has already shown to outperform state-of-the-art alternatives (Mukhopadhyay et al., 2016). We present the results in Fig.1, which shows that hierarchical survival models result in significantly improved likelihood over standard survival models. We also present the results of our severity prediction approach in Table1. We find an average accuracy of about 66%, a reasonable performance on a 3-class classification problem. To delve more deeply, note that in emergency response settings, incorrectly predicting high severity incidents is more costly than overestimating severity. We can see that the model rarely underestimates severity.

To evaluate the dynamic dispatch approach, we compared our approach with the existing approach followed by the actual fire department of our metropolitan area, which is in charge of dispatching ambulances to accidents. This baseline policy involves sending the closest available responder to an incident. To ensure the approach works under different levels of stress (the ratio of the frequency of incidents and the number of responders), we varied the number of responders in our experiments. We summarize the results in Table 2, which shows that at lower levels of stress, the MCTS approach and the baseline policy agree in most of the cases, but as the system is stressed, the MCTS approach impacts a significant number of incidents and saves vital response time. We highlight that this is a significant improvement, since the incidents served often involve life-and-death scenarios.

6 Deployment

In order for our work on emergency response to have meaningful impact, we created a tool (to be made open-source soon) for the fire and police departments of Nashville. The tool has already been deployed at the fire department, and we are working with the police department to deploy it.

7 Conclusion

In this paper, we summarized our work towards creating a systematic approach to ensure timely response to traffic accidents. We proposed learning a joint probability distribution over incident arrival and severity of incidents to tackle incident predictions heterogeneously based on priorities. Further, we proposed a novel hierarchical clustering approach to learn the spatial granularity at which such forecasting models should be learned. Finally, we formulated an MCTS based approach to tackle the dynamic dispatch of responders, that can use the incident prediction model to compute dispatch decisions quickly. Our experimental results demonstrate that our approaches offer significantly better performance than existing state-of-the-art approaches and policies followed in practice.

References

Williams Ackaah and Mohammed Salifu. Crash prediction model for two-lane rural highways in the ashanti region of ghana. *IATSS research*, 35(1):34–40, 2011.

Kevin Balke, Nadeem Chaudhary, Chi-Lueng Chu, Shamanth Kuchangi, Paul Nelson, Praprut Songchitruksa, Dvahg Swaroop, and Vipin Tyagi. Dynamic traffic flow modeling for incident detection and short-term congestion prediction: Year 1 progress report. *Texas Transportation Institute, Technical Report No. FHWAlTX-06/0–4946–1*, 2005.

Dankmar Böhning. Multinomial logistic regression algorithm. *Annals of the Institute of Statistical Mathematics*, 44(1):197–200, 1992.

David Roxbee Cox and David Oakes. *Analysis of survival data*, volume 21. CRC Press, 1984.

Wilpen Gorr, Andreas Olligschlaeger, and Yvonne Thompson. Short-term forecasting of crime. *International Journal of Forecasting*, 19(4):579–594, 2003.

Stephen C Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967.

Sean K Keneally, Matthew J Robbins, and Brian J Lunday. A markov decision process model for the optimal dispatch of military medical evacuation assets. *Health care management science*, 19 (2):111–129, 2016.

Leslie W Kennedy, Joel M Caplan, and Eric Piza. Risk clusters, hotspots, and spatial intelligence: risk terrain modeling as an algorithm for police resource allocation strategies. *Journal of Quantitative Criminology*, 27(3):339–362, 2011.

Ayan Mukhopadhyay, Chao Zhang, Yevgeniy Vorobeychik, Milind Tambe, Kenneth Pence, and Paul Speer. Optimal allocation of police patrol resources using a continuous-time crime model. In *Conference on Decision and Game Theory for Security*, 2016.

Ayan Mukhopadhyay, Yevgeniy Vorobeychik, Abhishek Dubey, and Gautam Biswas. Prioritized allocation of emergency responders based on a continuous-time incident prediction model. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pp. 168–177. International Foundation for Autonomous Agents and Multiagent Systems, 2017.

Ayan Mukhopadhyay, Zilin Wang, and Yevgeniy Vorobeychik. A decision theoretic framework for emergency responder dispatch. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 588–596. International Foundation for Autonomous Agents and Multiagent Systems, 2018.

Mark Poch and Fred Mannering. Negative binomial analysis of intersection-accident frequencies. *Journal of Transportation Engineering*, 122(2):105–113, 1996.

Venkataraman Shankar, Fred Mannering, and Woodrow Barfield. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Analysis & Prevention*, 27 (3):371–389, 1995.

Praprut Songchitruksa and Kevin Balke. Assessing weather, environment, and loop data for real-time freeway incident prediction. *Transportation Research Record: Journal of the Transportation Research Board*, (1959):105–113, 2006.

Paul W. Speer, D. M. Gorman, Erich W. Labouvie, and Mark J. Ontkush. Violent crime and alcohol availability: Relationships in an urban community. *Journal of Public Health Policy*, 19(3):303–318, 1998.

Chao Zhang, Arunesh Sinha, and Milind Tambe. Keeping pace with criminals: Designing patrol allocation against adaptive opportunistic criminals. In *International Conference on Autonomous Agents and Multiagent Systems*, pp. 1351–1359, 2015.

Chao Zhang, Victor Bucarey, Ayan Mukhopadhyay, Arunesh Sinha, Yundi Qian, Yevgeniy Vorobeychik, and Milind Tambe. Using abstractions to solve opportunistic crime security games at scale. In *International Conference on Autonomous Agents and Multiagent Systems*, 2016. to appear.