

SOS-EW: System for Overdose Spike Early Warning using Drug Mover’s Distance-based Hawkes Processes

Wen-Hao Chiang^{1†}, Baichuan Yuan^{2†}, Hao Li², Bao Wang², Andrea Bertozzi²,
Jeremy Carter¹, Brad Ray¹, and George Mohler^{1‡}

¹ Indiana University - Purdue University Indianapolis, USA

{chiangwe,carterjg,bradray,gmohler}@iupui.edu

² University of California Los Angeles, USA

{byuan,lihao0809,wangbao,bertozzi}@math.ucla.edu

Abstract. Opioid addictions and overdoses have increased across the U.S. and internationally over the past decade. In urban environments overdoses cluster in space and time, with 50% of overdoses occurring in less than 5% of the city and dozens of calls for emergency medical services being made within 48 hour periods. In this work, we introduce a system for early detection of opioid overdose clusters based upon the toxicology report of an initial event. We first use drug SMILES, one hot encoded molecular substructures, to generate a bag of drug vectors corresponding to each overdose (overdoses are often characterized by multiple drugs taken at the same time). We then use spectral clustering to generate overdose categories and estimate multivariate Hawkes processes for the space-time intensity of overdoses following an initial event. As the productivity parameter of the process depends on the overdose category, this allows us to estimate the magnitude of an overdose spike based on the substances present (e.g. fentanyl leads to more subsequent overdoses compared to Oxycontin). We validate the model using opioid overdose deaths in Indianapolis and show that the model outperforms several recently introduced Hawkes-Topic models based on Dirichlet processes. Our system could be used in combination with drug test strips to alert drug using populations of risky batches on the market or to more efficiently allocate naloxone to users and health/social workers.

Keywords: Opioid Overdose · Hawkes Process · Embedding · Spectral Clustering · Topic Model · Drug Mover’s Distance

1 Introduction

The United States is experiencing an overdose epidemic with more than a half million drug overdose deaths since 2,000 and over 70,000 drug overdose deaths

[†]Both authors contributed equally to this work.

[‡]Corresponding author.

in 2017 alone [43]. A majority of these deaths have been opioid-related overdoses; however, the role of opioids has varied dramatically across three waves of the epidemic, each resulting in increasing death rates [2]. This first wave began in the 1990s and was driven by prescription opioid-related deaths [3]. The reduced availability of these prescription medications is said to have resulted in the second wave of the epidemic, which began in 2010, and was driving by heroin-related deaths [3, 41, 45]. The third wave started in 2013 and has largely been driven by illicit fentanyl, a synthetic opioid that is 50 to 100 times more potent than morphine [12]. National estimates suggest that in 2016 nearly half of opioid-related deaths contained fentanyl [16], and there is evidence showing that fentanyl is being mixed into heroin and cocaine which is likely contributing overdose deaths involving these substances [17, 29]. Recent research has also shown that opioid overdoses cluster in space and time, where over half of opioid overdose deaths may occur in less than 5% of a city [1].

As the overdose epidemic has progressed researchers and policy makers have revealed shortcomings in official data sources, namely vital records data. One limitation is that vital records data rely on the International Classification of Diseases, 10th Revision (ICD-10) codes which do not record the specific substances related to an overdose fatality [6, 15, 27, 48]; for example, there is no ICD code for fentanyl. Another limitation though has been the undercounting of opioid-related fatalities as 20 to 35 percent of drug overdose deaths are unspecified [42], meaning no substance was indicated as a primary or contributing cause of death. Moreover, rates of undercounting vary geographically as they are the result of state policies for death investigation procedures [42]. While researchers have developed measures to adjust for these limitations [42], better data collection systems are being implemented [47], and state policies are changing [11], it remains clear that we lacked sufficient data to quickly detect and identify the substances driving this overdose epidemic at the national level. However, one source of local information that can be used to address these gaps are toxicology results collected as part of an overdose death investigation. In the present study, we use a robust toxicology dataset from Marion County, Indiana [Indianapolis] that were collected as part of the CDCs Prevention for States funding initiative [28, 37, 38]. In the toxicology dataset, we observe spatio-temporal clustering patterns [1]. These shift patterns demonstrate that overdose events concentrate within micro places in a short time window and shift through time, which motivates our present work.

Our goal in the present work is to develop a statistical framework for modeling and prediction of opioid overdose clusters in space and time, leveraging information provided in the toxicology report of the initial overdose in the cluster. An overview of our proposed system, SOS-EW, is given in Figure 1. We use a Hawkes process to model overdoses as a branching process. Each event may trigger offspring events nearby in space and time. The branching ratio of the process, determining the average number of offspring, depends on the drugs contained in the toxicology report of each parent event in the branching process. To reduce the dimension, we use spectral clustering with earth mover’s distance

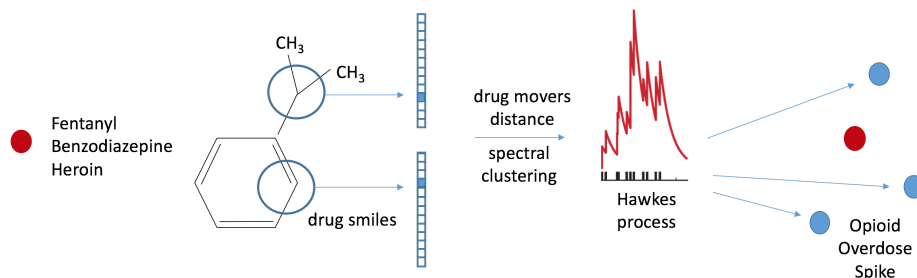


Fig. 1: Overview of the SOS – EW system for early warning of opioid spikes. The initial overdose toxicology report shows fentanyl, benzodiazepine, and heroin present. Each drug is vectorized using SMILES and the event is assigned an overdose category using spectral clustering based on earth mover’s distance of the drug vectors (“drug mover’s distance”). The increase in the intensity of the Hawkes process is determined by the category and allows for the prediction of an opioid overdose spike, those events triggered in the branching process by the initial overdose.

on bags of drug SMILES [14] vectors corresponding to each toxicology report (we refer to in the rest of the paper as “drug mover’s distance”). The resulting method outperforms existing Dirichlet-based Hawkes topic models in the task of early warning of opioid overdose spikes (clusters) based on an initial event and its toxicology report.

The outline of the paper is as follows. In Section 2, we review related work on point process models of event clustering in urban environments and topic point processes for event data with high dimensional marks. In Section 2, we then present the details on our proposed method including clustering of the toxicology reports and the spatio-temporal Hawkes process. In Section 3, we provide details for several baseline models we use to benchmark SOS-EW and in Section 4 we present the results of our numerical experiments. We then discuss the implications of our results for practice in Section 5.

2 Method

SOS – EW system for early warning of opioid spikes is mainly comprised of the following two components:

1. Overdose category clustering;
2. Marked spatio-temporal Hawkes process kernel estimation and simulation.

Given a toxicology report, overdose events are first clustered into several overdose categories through spectral clustering [34]. The distance between overdose events in spectral clustering is measured based on earth mover’s distance [40] of drug vectors which characterize drug’s two-dimensional molecular structure (in particular we use SMILES, one hot encoded molecular substructures [14]). Drug overdose events over continuous time are then modeled through spatio-temporal

Hawkes processes. We estimate the productivity for each category based on historical overdose events and simulate future events to generate a short ranked list of hotspots containing overdose spikes. Figure 1 presents an overview of the SOS – EW system. In Section 2.1, we review related work on spatial self-exciting point processes, along with topic point processes. In Section 2.2, we present the details of our method for overdose event clustering and we introduce our approach for the estimation and simulation of spatio-temporal Hawkes processes in Section 2.3.

2.1 Related Work

Self-exciting (Hawkes) point processes have been used to model space-time clustering in urban crime patterns [32] and Hawkes process-based learning to rank algorithms were recently a top-performing solution in the 2017 NIJ Crime Forecasting competition [31], which focused on ranking the top crime hotspots in a city according to short-term crime risk. Other point process models, for example, log-Gaussian Cox Processes, can model spatial diffusion of events and have also proven accurate at modeling crime hotspots [8, 44]. More recently, self-exciting point processes have been used to model clustering in emergency call data [25]. In more extreme security settings spatio-temporal point process models for event prediction have been applied to conflict [51] and terrorism [10] datasets.

In the above studies, the models either only used as input the spatial coordinates and time of the events, or in some cases an additional low-dimensional (< 10) event category. However, event data often is accompanied by a high dimensional mark, for example, text data, imagery, sensor data, or in our case a 133 dimensional vector indicating drugs in a toxicology screen.

Recent work in the machine learning and information retrieval literature has focused on extending temporal and network-based Hawkes processes to handle text information in the events [21, 4]. Dirichlet Hawkes processes [4, 49] have been introduced for this purpose, where document clustering is jointly learned with a temporal Hawkes process. In the network setting, Hawkes processes have been used to model coupled information and event diffusion on networks [5]. However, these studies have not dealt with spatio-temporal data, which is critical in studying the spread of opioid overdoses.

Our work offers several contributions to the above-related literature. First, we investigate the applicability of existing Hawkes-topic models in the spatial setting and then we improve upon the accuracy of these models in several prediction tasks related to early warning of opioid overdose clusters. Second, we introduce a novel clustering method for drug overdoses based upon drug mover’s distance. Related to word mover’s distance [20] that has shown higher coherence than LDA based topic models, we believe our drug mover’s distance-based spectral clustering may be useful in a variety of applications where sets of molecules need to be compared and clustered.

2.2 Overdose Categorization

To categorize overdose events through clustering, we first measure the similarity between two overdose events in terms of a “distance” based on the drugs involved in the events. Each event consists of a mark indicating one or more drug substances found in the victim’s system. We denote an event i containing m drug substances as $\mathbf{E}_i = \{\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_m\}$, where drug m is denoted as \mathbf{d}_m . Each drug is represented by a set of 2D substructures, i.e., $\mathbf{d} \in \{s_1, s_2, \dots, s_k\}$, where the substructure k is denoted as s_k . The distance between each drug event is then calculated by earth mover’s distance [40].

Earth mover’s distance (EMD) is a metric to measure a distance between two distributions. EMD is based on the minimal cost that must be paid to move one distribution into the other. Given two events, \mathbf{E}_p and \mathbf{E}_q , with m and n drugs, respectively, we want to find a transportation flow $\mathbf{F} \in \mathbb{Z}_2^{m \times n}$, where $\mathbb{Z}_2 = \{0, 1\}$, that minimizes the overall cost:

$$\begin{aligned} & \min \sum_{i=1}^m \sum_{j=1}^n F_{ij} C_{ij}, \\ \text{subject to } & \sum_{j=1}^n F_{ij} \leq 1 \quad 1 \leq i \leq m, \\ & \sum_{i=1}^m F_{ij} \leq 1 \quad 1 \leq j \leq n, \\ & \sum_{i=1}^m \sum_{j=1}^n F_{ij} \leq \min(m, n). \end{aligned} \tag{1}$$

C_{ij} is the cost for moving \mathbf{d}_i to \mathbf{d}_j . We define such cost as Jaccard distance [23] and it can be calculated by dividing the difference of the sizes of the union and the intersection of two sets of substructure by the size of the union:

$$C_{ij} = \frac{|\mathbf{d}_i \cup \mathbf{d}_j| - |\mathbf{d}_i \cap \mathbf{d}_j|}{|\mathbf{d}_i \cup \mathbf{d}_j|}. \tag{2}$$

Such an optimization problem 1 can be further solved through the transportation simplex method [26]. Once the optimal transportation flow is found, the EMD between event \mathbf{E}_p and \mathbf{E}_q is defined as the resulting overall cost normalized by the total transportation flow:

$$\text{EMD}(\mathbf{E}_p, \mathbf{E}_q) = \frac{\sum_{i=1}^m \sum_{j=1}^n F_{ij} C_{ij}}{\sum_{i=1}^m \sum_{j=1}^n F_{ij}}. \tag{3}$$

After calculating the EMD between each overdose events, we then construct a similarity matrix (i.e., adjacency matrix). The similarity between \mathbf{E}_p and \mathbf{E}_q is calculated using a radial basis function kernel, i.e., $\exp\left(\frac{-\text{EMD}(\mathbf{E}_p, \mathbf{E}_q)^2}{2\epsilon^2}\right)$. To categorize drug overdose events into different clusters, we apply spectral

clustering [46] on all the events. The spectral clustering algorithm takes in the adjacency matrix and uses the eigenvalues and eigenvectors from the adjacency matrix of the events to perform dimensionality reduction before clustering. Each overdose event is then assigned to a category u . Therefore, each event $\mathbf{E}_i = (t_i, x_i, y_i, u_i)$ consists of four pieces of information: t_i is the timestamp of the date of death (D.o.D) of the victim; x_i and y_i are the latitude and longitude of where the victim is found; and u_i is the drug overdose category. Our approach to overdose categorization can seamlessly integrate the molecular substructure similarities across different drug overdoses and produce more pharmacokinetic-aware categories. Figure 2 shows the top 5 overdose categories, along with the most frequent drugs in each category, computed using our drug mover’s distance-based spectral clustering approach.

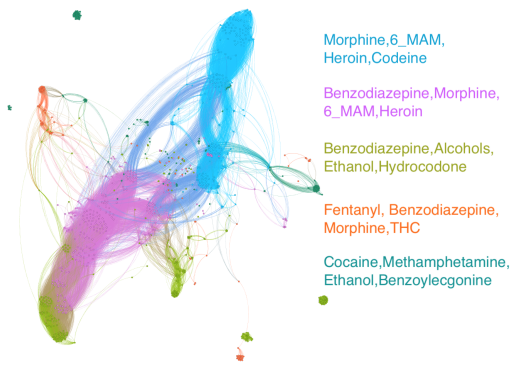


Fig. 2: Gephi visualization of graph used for spectral clustering (edges correspond to adjacency weight greater than .9) and the most frequent drugs in each of the 5 overdose categories.

2.3 Spatio-temporal Hawkes Process

Clustering in space-time drug overdose event data may occur for a variety of reasons; for example, an increase in the prevalence of a new street drug may appear in a neighborhood leading to a spike in overdoses; or a particular batch of drugs may contain a higher than usual amount of a dangerous substance, for example, fentanyl. Motivated by the observed clustering of overdose data [1], we further characterize drug overdose events through a cross-exciting spatio-temporal Hawkes process [50] that models the contagiousness of events across overdose categories (computed using DMD spectral clustering).

Given a drug overdose sequence $\{\mathbf{E}_i\}_i^N$ with N events, we characterize a multivariate spatio-temporal Hawkes process through the following conditional intensity function for each category u :

$$\lambda_u(x, y, t) = \mu_u(x, y) + \sum_{t > t_i} K_{u_i u} g(x - x_i, y - y_i, t - t_i). \quad (4)$$

In Equation (4), the background rate $\mu_u(x, y)$ for each category is assumed to be a constant in time, while inhomogeneous in space. The historical events increase the likelihood of the near-future events through the spatio-temporal triggering density function g . $\mathbf{K}(u_i, u_j) = K_{u_i, u_j}$ is the productivity (or triggering) matrix

to quantify the self or cross-exciting impact of the events associated with category u_i on the subsequent events in the category u_j . Specifically, K_{u_i, u_j} denotes the mean number of events in process u_j that are triggered by each event in the process u_i .

We introduce a parametric form of the triggering density with an exponential function in time and a Gaussian density in space. These choices allow for a weighted sample mean estimation of the parameters in the maximization (M) step of expectation-maximization (EM) based maximum likelihood estimation (MLE) [30].

Our kernel density-based background rate and triggering density function take the following form:

$$\begin{aligned}\mu_u(x, y) &= \sum_{i=1}^N \frac{\beta_{u_i u}}{2\pi\eta^2 T_{\text{span}}} \times \exp\left(-\frac{(x-x_i)^2 + (y-y_i)^2}{2\eta^2}\right), \\ g(x, y, t) &= \omega \exp(-\omega t) \times \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right),\end{aligned}\tag{5}$$

where T_{span} denotes the time spanned through the whole training dataset; $\beta_{u_i u}$ measures the extent to which events in process u_i contribute to the background rate in the process u_j ; ω controls how fast the rate $\lambda_u(x, y, t)$ returns to its baseline level $\mu_u(x, y)$ after an event occurs; and η and σ dictate the spreading scale of the triggering effect in space.

We perform the M step of the EM-type algorithm following the framework of Algorithm 1 in [50] to estimate the parameters. We use the ‘‘optimal’’ parameters from the previous M step to update the latent variables and alternately iterate E and M step. After parameter estimation, we utilize the branching structures [52] of self-exciting point processes to simulate self and cross-exciting events (See Algorithm 3 in [50]) for the next T days for 1,000 times. The simulated events are denoted as $\hat{\mathbf{E}} = (\hat{t}, \hat{x}, \hat{y}, \hat{u})$.

To make recommendations for early warning of overdose spikes, we generate a short ranked list of hotspots in the domain of interest. We first partition the domain of interest into $N \times N$ fine-grained grid cells by dividing the latitude and longitude span into N parts with equal length. Based on the latitude and longitude (i.e., \hat{x} and \hat{y}) from the simulated events, we calculate the average number of the simulated events for each grid cell from 1,000 repeated simulations. We denote the average number of the simulated events in the i^{th} and j^{th} grid in terms of latitude and longitude as $\chi(i, j)$. Finally, we sort the grids according to the average number of the simulated events in descending order and retain the top- N grids as the recommended short ranked list.

3 Comparison Methods

We compare our model with several state-of-the-art methods including the following: Non-parametric temporal Hawkes Processes; Spatio-temporal univariate

Hawkes Processes; and Dirichlet Hawkes Processes that learn the category assignment while estimating the intensity function. None of the existing point process methods jointly learn the spatio-temporal self and cross-exciting density together for different overdose categories based on the drug substance structure; methods either exist for spatio-temporal point processes or for time-only topic point processes, but not both. The details of our implementation for these baseline methods are presented in the following sections.

Non-parametric Hawkes Processes: `SimpHP` [24]

Our first baseline model utilizes a non-parametric Hawkes process [24] that takes a series of time stamps and then uses a penalized MLE to simultaneously estimate the background rate $\mu(t)$ and triggering kernel $g(t)$ without prior knowledge of their form. Given a set of overdose events, we partition the training dataset into $N \times N$ subsets according to which fine-grained grid cell they belong (as defined in Section 2.3). Each subset of events corresponds to an independent Hawkes process. Once the estimation is done, the simulation for the next T interval is done through thinning [35]. Each Hawkes process is first simulated 1,000 times and the average number of simulated events is calculated. The top- N grids cell with the largest average number of simulated events is recommended. We denote this baseline method as `SimpHP`.

Spatial-temporal Hawkes Processes: `SpatHP` [50]

We compare `SOS – EW` with a sub-model that only uses the geolocation and time stamps for estimation, without clustering events into different categories (univariate). The model estimation and recommendation follow the same framework in Section 2.3. Such a baseline model is denoted as `SpatHP`.

Dirichlet Hawkes Processes: `TopicHP` [4]

The Dirichlet Hawkes process [4] is a random process which takes into account both text information from documents and temporal dynamics of their arrival pattern to cluster the document streams. The model is estimated through an online inference algorithm that jointly learns the pattern of the clusters and the parameters of the Hawkes process for each cluster. To adapt to this model, we view each overdose event as a document and each drug as a word. Spatial information is integrated into each event by adding a grid cell index as an additional word. After model estimation, we then use thinning [35] based simulation and average the number of events for each grid cell and topic (over 1,000 simulations) to generate a recommendation of the top- N grid cells with the most number of simulated events.

4 Experiments

4.1 Data

We analyze a toxicology dataset from Marion County, Indiana that was collected as part of the CDCs Prevention for States funding initiative [28, 37, 38].

The dataset contains toxicology reports of 1,489 overdose death events in Indianapolis, Indiana, U.S.A. from 2010 to 2016. Each overdose event includes the date of death (D.o.D) of the victim and the geolocation (latitude and longitude) of where the victim was found. In addition, every event also contains forensic toxicology testing results that screen for 164 drug substances. A binary indicator represents whether a specific drug substance was found in the victim’s body. In our analysis, we restricted to a subset of 133 drugs whose 2D chemical structure representation can be found in the chemical molecules database, Pubchem [19], for further feature generation. We also restricted our analysis to the 1,425 overdose events that include geolocation information and occurred within the city of Indianapolis boundary, where the latitude ranges from 39°37’58.8”N to 39°55’30.3”N and the longitude ranges from 87°06’41.1”W to 85°56’18.7”W. Table 1 presents the statics of the pruned dataset. Figure 4 presents the example of some overdose events.

Table 1: Statistics on Overdoes Event

start date	01-14-2010
end date	12-30-2016
#event	1,425
$\overline{\#d}$	6.3698
$\overline{\#ent_d}$	68.2481

In this table, “#event” represents the number of events in the toxicology report; “ $\overline{\#d}$ ” represents the average number of drug substances in each event; and “ $\overline{\#ent_d}$ ” is the average number of events involved in each drug.

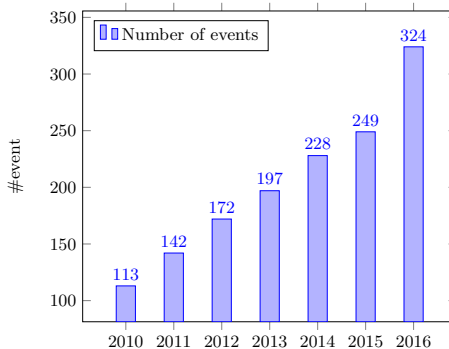


Fig. 3: Number of Events per Year

Features for each drug are extracted by identifying its molecular substructure fingerprints. Specifically, each drug is represented by a set of substructures s (i.g., $\mathbf{d} \in \{s_1, s_2, \dots, s_k\}$). We further used RDKit [22], open-source software that allows us to search the substructures based on 2D chemical structures representation, to generate a feature vector of dimension 1,024 for each drug. The pruned dataset is then used for model evaluation.

4.2 Evaluation Protocols and Metrics

The domain of Indianapolis covered by the pruned dataset is first partitioned into $N \times N$ grid cells by dividing the latitude and longitude span into N parts with equal length in each direction. For each time interval t^{th} , we recommend a ranked list of grid cells based on how likely those grid cells are to have overdose events in the near future, using the history of the process up to t^{th} time

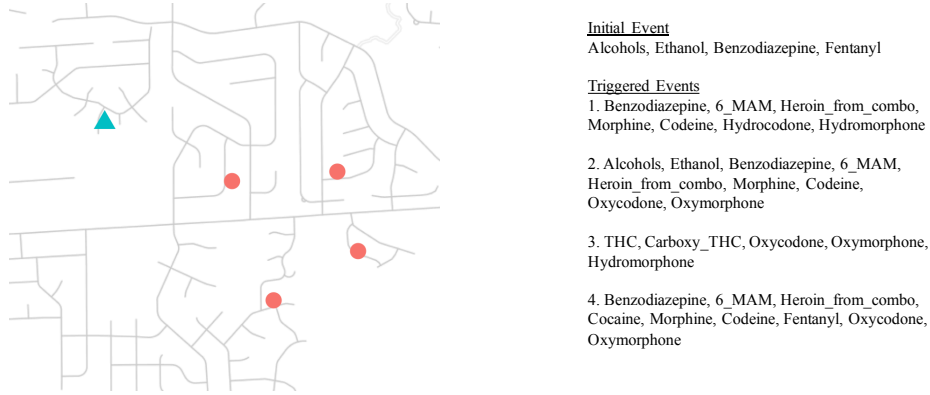


Fig. 4: An illustration for an initial event and its triggered events in one of the categories (i.e., one of the Hawks processes). The initial overdose event marked in triangle symbol consists of four drug substances and it triggered four neighboring events consisting of different number of drug substances respectively.

interval. The performance is then evaluated through walk forward optimization [30]. Specifically, in our experimental setting, we first train our model over a fixed amount of the most recent historical events, which is 255 events (the number of events in the first two years). We then test the model on the next time period starting on 01/01/2012 and report the performance. Finally, the overall performance is the average from all the time periods that we have tested.

At each time interval T , models recommend a ranked list of size K for potential overdose events. In our experimental setting, we set T as 5 days and partition the domain of interest into 100×100 grid cells. The time window is consistent with the time scale on which police and health services can respond and the grid cell size is similar to those used in field trials of predictive policing [33]. The ranking performance is evaluated through normalized discounted cumulative gain at K ($NDCG@K$), which is a measurement of ranking quality and commonly used in information retrieval. $NDCG@K$ is calculated by normalizing discounted cumulative gain (denoted as $DCG@K$) with ideal $DCG@K$ (denoted as $IDCG@K$). The definition are as follows:

$$DCG@K = \sum_{i=1}^K \frac{2^{rel_i} - 1}{\log_2(i + 1)}, \quad NDCG@K = \frac{DCG@K}{IDCG@K}, \quad (6)$$

where rel_i is the i^{th} relevance value of the i^{th} grid in the ranked list \mathbf{r} ; $IDCG@K$ is the ideal $DCG@K$ when the ranked list \mathbf{r} is perfectly ranked based on its relevance values; and then we define relevance value rel_i as the number of overdose events that happen in the i^{th} grid cell in the ranked list \mathbf{r} and t^{th} time interval.

In a certain time interval, a spike of overdose events may occur. To evaluate our model's ability to forecast future spikes, we first define an event spike at the

t^{th} time interval in a grid cell (i, j) as follows: If the total number of events in the neighboring cells between the $(t - 1)^{\text{th}}$ and $(t + 1)^{\text{th}}$ intervals is more than a threshold, then we consider that there is an event spike in (i, j) grid within these time intervals. The set of grids with an event spike is defined as the following:

$$\mathbf{s} = \{(i, j, t) \mid \sum_{\hat{i}=i-w}^{i+w} \sum_{\hat{j}=j-w}^{j+w} \sum_{\hat{t}=t-1}^{t+1} \chi(\hat{i}, \hat{j}, \hat{t}) \geq \xi\} \quad (7)$$

where $\chi(i, j, t)$ is the count of events in grid cell (i, j) and time interval t , w is the spatial window size defining how many neighboring cells we should consider; and ξ is the spike threshold: we set $\xi = 2$ (for $w = 2$ and 4) and $\xi = 10$ (for $w = 10$ and 15) in our evaluation setting. With larger w and ξ , more grid cells will be considered to have an event spike. We choose w and ξ to ensure a reasonable amount of event spikes for further evaluation while an event spike includes a huge amount of overdose events. We adopt modified reciprocal hit rank [36], precision and recall at different ranked list size K , denoted as $MRHR$, $Prec$, and Rec to evaluate the performance. $MRHR$ is a modified version of average reciprocal hit rank (ARHR), which is feasible for ranking evaluations where there are multiples relevant items (i.e., multiple spikes events), and it is calculated as the following:

$$MRHR = \frac{1}{|\mathbf{s}|} \sum_{i=1}^K \left(\frac{hit_i}{rank_i} \right), \quad (8)$$

$$\text{where } hit_i = \begin{cases} 1 & \text{if } r_i \in \mathbf{s} \\ 0 & \text{if } r_i \notin \mathbf{s} \end{cases}, rank_i = \begin{cases} rank_{i-1} & \text{if } hit_{i-1} = 1 \\ rank_{i-1} + 1 & \text{if } hit_{i-1} = 0, \end{cases}$$

where each hit is rewarded based on its position in the ranked list. $Prec$ and Rec are commonly used to evaluate the performance in recommendation system. $Prec$ evaluates how precisely the model can predict for future spike events while Rec measures the ability of retrieving spikes. We also evaluate average precision, denoted as APC , to account for both precision and recall without choosing K :

$$Prec = \frac{|\mathbf{s} \cap \mathbf{r}|}{|\mathbf{r}|}, \quad Rec = \frac{|\mathbf{s} \cap \mathbf{r}|}{|\mathbf{s}|}, \quad APC = \frac{\sum_{\{k:1, \dots, K | r_k \in \mathbf{s}\}} Prec@k}{|\mathbf{s} \cap \mathbf{r}|}. \quad (9)$$

4.3 Experimental Results

Table 2 presents the overall performances on overdose spikes prediction under different breadth definition of spike events (Equation 7). Our **SOS – EW** system consistently outperforms other baseline methods by a large margin in terms of $MRHR$, $Prec$, Rec , and APC . $MRHR$ is used to evaluate the ranking quality while $Prec$, Rec , and APC evaluate the retrieval for events spikes. This shows our proposed method not only can successfully recommend the regions with potential event spikes considering the ranking position but it can also precisely generate a short ranked list for those event spikes precisely. As the ranking list size increases, performances of $MRHR$ remains similar after $MRHR$

Table 2: Overall Performance on Different Spike Window Size w

w	mdl	<i>MRHR</i>			<i>Prec</i>			<i>Rec</i>			<i>APC</i>
		@1%	@3%	@5%	@1%	@3%	@5%	@1%	@3%	@5%	
2	SOS – EW	.0172	.0188	.0191	.0277	.0174	.0124	.2534	.5038	.6110	.0451
	SimpHP	.0088	.0089	.0090	.0034	.0019	.0018	.0289	.0496	.0842	.0103
	SpatHP	.0017	.0021	.0023	.0036	.0029	.0025	.0375	.1129	.1590	.0043
	TopicHP	.0080	.0082	.0082	.0130	.0085	.0055	.1258	.2386	.2748	.0125
4	SOS – EW	.0218	.0236	.0240	.1091	.0724	.0555	.2690	.5196	.6699	.1187
	SimpHP	.0059	.0060	.0060	.0126	.0070	.0060	.0279	.0475	.0660	.0118
	SpatHP	.0016	.0018	.0020	.0114	.0111	.0112	.0274	.0784	.1294	.0102
	TopicHP	.0082	.0089	.0090	.0474	.0315	.0205	.1089	.2226	.2434	.0250
10	SOS – EW	.0219	.0254	.0259	.2000	.1526	.1229	.1911	.6734	.8554	.1851
	SimpHP	.0029	.0031	.0032	.0189	.0119	.0111	.0128	.0662	.0854	.0125
	SpatHP	.0003	.0007	.0007	.0167	.0181	.0193	.0109	.0691	.1025	.0151
	TopicHP	.0096	.0097	.0098	.0856	.0548	.0360	.1974	.2868	.3121	.0335
15	SOS – EW	.0620	.0718	.0726	.4387	.3275	.2549	.3234	.6470	.8025	.3644
	SimpHP	.0070	.0072	.0072	.0445	.0291	.0254	.0336	.0526	.0676	.0301
	SpatHP	.0016	.0019	.0020	.0516	.0506	.0503	.0220	.0746	.1119	.0387
	TopicHP	.0083	.0091	.0092	.1739	.1212	.0802	.1035	.2081	.2267	.0622

The column “mdl” corresponds to different models. The best overall performance is **bold**.

@3%; *Prec* decreases while *Rec* increases due to the natural trade-off between these two metrics. SOS – EW estimates the model parameters and makes predictions specifically for different overdose categories compared to SpatHP which estimates the same parameters for all event data aggregated together. This indicates that strategically clustering overdose events based on the drug molecular structure can achieve better performances than the model which solely relies on spatio-temporal information. TopicHP jointly learns the clustering structure and model parameters and it can be viewed as a competitive baseline. However, only the drug distribution in each event is taken into account and drugs’ chemical structure is not included in the model training. This may explain why TopicHP falls short of spikes recommendation metrics compared to SOS – EW but is still better than other baselines. Overall, SOS – EW leverages the information from geo-locations, event triggering dynamics, and drugs’ high level physical and chemical properties based on 2D structures altogether and precisely makes the recommendations for future overdose event spikes.

Table 3 presents the overall model performances on *NDCG* when only evaluating each models ability to rank grid cells (equivalent to $\omega = 0$). The ranking quality increases as the size of the ranked list grows larger (i.e., from 1% to 5%). Our proposed SOS – EW outperforms the second best baseline method TopicHP by 12.27% at *NDCG* @5%, however, TopicHP is the top performing model

Table 3: Overall *NDCG* Performance

mdl	@1%	@3%	@5%
SOS – EW	.0322	.0637	.0842
SimpHP	.0191	.0236	.0265
SpatHP	.0092	.0179	.0262
TopicHP	.0492	.0733	.0750

The column “mdl” corresponds to different models. The best overall performance is **bold**.

at 1% and 3%. These results indicate that for the very highest risk cells, spatial diffusion may play less of a role and `TopicHP` and `SOS – EW` have similar performance. However, for flagging neighborhoods instead of individual cells for spikes, `SOS – EW` is superior due to its ability to model spatial diffusion of risk.

5 Implications for Practice

Results from our proposed system can be translated into more effective social service delivery and intervention programming. When flagging the top 1% of predicted spikes defined in neighborhoods of approximately $O(1km^2)$ in size and 5 days in length, the method captures around 25% of opioid spikes. These spatial and temporal scales are similar in size to those used in predictive policing [33] and by efficiently predicting the geographic diffusion of opioid-related toxins (such as fentanyl), social service programs and first responders can develop dynamic programs to best target areas where people face the highest risk of overdose. Further research is needed to verify whether or not the results found in Indianapolis in this study extend to other cities and rural areas.

Studies have revealed polydrug patterns whereby fentanyl is being detected alongside cocaine and methamphetamines, which is contributing to overdose deaths involving these substances [17, 29]. Given the nature of this supply-side poisoning among illicit drugs, the most feasible approach may be to empower and provide persons who used drugs with the ability to test these substances. Drug testing technologies (i.e., fentanyl test strips) allow drug users to understand whether the drugs they use are contaminated with lethal substances, such as fentanyl, which can allow them to adjust behaviors and prevent a potentially fatal overdose [13]. Furthermore, the average of dispatch and response time for emergency medical services (EMS) personnel to arrive when an overdose event is reported is seven minutes on average [‡] and the time to results of many drug test strips is usually less than minutes nowadays. Therefore, our system can make a prediction in a short time so that the health/social workers can react accordingly. More research is needed to extend the method in this paper, that utilizes toxicology reports, to the application of drug test strips and other drug testing tools (which may not be as thorough as a coroner’s report).

Public health services can deploy syringe services, such as retractable syringes or exchange programs, that have been shown to reduce fatal opioid use [18, 9]. From a policing perspective, officers can be equipped with nasal naloxone (or Narcan) within high-risk opioid locations to reduce the likelihood of death from an opioid overdose [7, 39].

Acknowledgements

[‡]<https://www.medicalnewsbulletin.com/response-time-emergency-medical-services/>

This work was supported by NIJ grant 2018-R2-CX-0013, NSF grants DMS-1737770, SCC-1737585, ATD-1737996, and DARPA grant FA8750-18-2-0066.

References

1. Carter, J.G., Mohler, G., Ray, B.: Spatial concentration of opioid overdose deaths in indianapolis: An application of the law of crime concentration at place to a public health epidemic. *Journal of Contemporary Criminal Justice* p. 1043986218803527 (2018)
2. Ciccarone, D.: Fentanyl in the us heroin supply: a rapidly changing risk environment. *International Journal of Drug Policy* **46**, 107–111 (2017)
3. Cicero, T.J., Ellis, M.S., Surratt, H.L., Kurtz, S.P.: The changing face of heroin use in the united states: a retrospective analysis of the past 50 years. *JAMA psychiatry* **71**(7), 821–826 (2014)
4. Du, N., Farajtabar, M., Ahmed, A., Smola, A.J., Song, L.: Dirichlet-hawkes processes with applications to clustering continuous-time document streams. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 219–228. ACM (2015)
5. Farajtabar, M., Wang, Y., Rodriguez, M.G., Li, S., Zha, H., Song, L.: Coevolve: A joint point process model for information diffusion and network co-evolution. In: *Advances in Neural Information Processing Systems*. pp. 1954–1962 (2015)
6. Fernandez, W., Hackman, H., Mckeown, L., Anderson, T., Hume, B.: Trends in opioid-related fatal overdoses in massachusetts, 1990–2003. *Journal of substance abuse treatment* **31**(2), 151–156 (2006)
7. Fisher, R., O'Donnell, D., Ray, B., Rusyniak, D.: Police officers can safely and effectively administer intranasal naloxone. *Prehospital Emergency Care* **20**(6), 675–680 (2016)
8. Flaxman, S., Chirico, M., Pereira, P., Loeffler, C.: Scalable high-resolution forecasting of sparse spatiotemporal events with kernel methods: a winning solution to the nij” real-time crime forecasting challenge”. *arXiv preprint arXiv:1801.02858* (2018)
9. Frost, M.C., Williams, E.C., Kingston, S., Banta-Green, C.J.: Interest in getting help to reduce or stop substance use among syringe exchange clients who use opioids. *Journal of addiction medicine* **12**(6), 428–434 (2018)
10. Gao, P., Guo, D., Liao, K., Webb, J.J., Cutter, S.L.: Early detection of terrorism outbreaks using prospective space–time scan statistics. *The Professional Geographer* **65**(4), 676–691 (2013)
11. Gilson, T.P.: The medical examiner’s role in addressing the opioid crisis. *US Att’y’s Bull.* **66**, 47 (2018)
12. Gladden, R.M.: Fentanyl law enforcement submissions and increases in synthetic opioid–involved overdose deaths27 states, 2013–2014. *MMWR. Morbidity and mortality weekly report* **65** (2016)
13. Glick, J.L., Christensen, T., Park, J.N., McKenzie, M., Green, T.C., Sherman, S.G.: Stakeholder perspectives on implementing fentanyl drug checking: Results from a multi-site study. *Drug and alcohol dependence* **194**, 527–532 (2019)
14. Hirohara, M., Saito, Y., Koda, Y., Sato, K., Sakakibara, Y.: Convolutional neural network based on smiles representation of compounds for detecting chemical motif. *BMC bioinformatics* **19**(19), 526 (2018)

15. Hoppe-Roberts, J.M., Lloyd, L.M., Chyka, P.A.: Poisoning mortality in the united states: comparison of national mortality statistics and poison control center reports. *Annals of emergency medicine* **35**(5), 440–448 (2000)
16. Jones, C.M., Einstein, E.B., Compton, W.M.: Changes in synthetic opioid involvement in drug overdose deaths in the united states, 2010-2016. *Jama* **319**(17), 1819–1821 (2018)
17. Kandel, D.B., Hu, M.C., Griesler, P., Wall, M.: Increases from 2002 to 2015 in prescription opioid overdose deaths in combination with other substances. *Drug and alcohol dependence* **178**, 501–511 (2017)
18. Karamouzian, M., Dohoo, C., Forsting, S., McNeil, R., Kerr, T., Lysyshyn, M.: Evaluation of a fentanyl drug checking service for clients of a supervised injection facility, vancouver, canada. *Harm reduction journal* **15**(1), 46 (2018)
19. Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A., et al.: Pubchem substance and compound databases. *Nucleic acids research* **44**(D1), D1202–D1213 (2015)
20. Kusner, M., Sun, Y., Kolkin, N., Weinberger, K.: From word embeddings to document distances. In: *International Conference on Machine Learning*. pp. 957–966 (2015)
21. Lai, E.L., Moyer, D., Yuan, B., Fox, E., Hunter, B., Bertozzi, A.L., Brantingham, P.J.: Topic time series analysis of microblogs. *IMA Journal of Applied Mathematics* **81**(3), 409–431 (2016)
22. Landrum, G.: Rdkit: open-source cheminformatics software (2016)
23. Levandowsky, M., Winter, D.: Distance between sets. *Nature* **234**(5323), 34 (1971)
24. Lewis, E., Mohler, G.: A nonparametric em algorithm for multiscale hawkes processes. preprint (2011)
25. Li, C., Song, Z., Wang, X.: Nonparametric method for modeling clustering phenomena in emergency calls under spatial-temporal self-exciting point processes. *IEEE Access* **7**, 24865–24876 (2019). <https://doi.org/10.1109/ACCESS.2019.2900340>
26. Lieberman, G.J., Hillier, F.: *Introduction to mathematical programming*. McGraw-Hill (1995)
27. Linakis, J.G., Frederick, K.A.: Poisoning deaths not reported to the regional poison control center. *Annals of emergency medicine* **22**(12), 1822–1828 (1993)
28. Lowder, E.M., Ray, B.R., Huynh, P., Ballew, A., Watson, D.P.: Identifying unreported opioid deaths through toxicology data and vital records linkage: Case study in marion county, indiana, 2011–2016. *American journal of public health* **108**(12), 1682–1687 (2018)
29. McCall Jones, C., Baldwin, G.T., Compton, W.M.: Recent increases in cocaine-related overdose deaths and the role of opioids. *American journal of public health* **107**(3), 430–432 (2017)
30. Mohler, G.: Marked point process hotspot maps for homicide and gun crime prediction in chicago. *International Journal of Forecasting* **30**(3), 491–497 (2014)
31. Mohler, G., Porter, M.D.: Rotational grid, pai-maximizing crime forecasts. *Statistical Analysis and Data Mining: The ASA Data Science Journal* **11**(5), 227–236 (2018)
32. Mohler, G.O., Short, M.B., Brantingham, P.J., Schoenberg, F.P., Tita, G.E.: Self-exciting point process modeling of crime. *Journal of the American Statistical Association* **106**(493), 100–108 (2011)
33. Mohler, G.O., Short, M.B., Malinowski, S., Johnson, M., Tita, G.E., Bertozzi, A.L., Brantingham, P.J.: Randomized controlled field trials of predictive policing. *Journal of the American statistical association* **110**(512), 1399–1411 (2015)

34. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: Analysis and an algorithm. In: *Advances in neural information processing systems*. pp. 849–856 (2002)
35. Ogata, Y.: On lewis’ simulation method for point processes. *IEEE Transactions on Information Theory* **27**(1), 23–31 (1981)
36. Peker, S., Kocyigit, A.: mrhr: a modified reciprocal hit rank metric for ranking evaluation of multiple preferences in top-n recommender systems. In: *International Conference on Artificial Intelligence: Methodology, Systems, and Applications*. pp. 320–329. Springer (2016)
37. Phalen, P., Ray, B., Watson, D.P., Huynh, P., Greene, M.S.: Fentanyl related overdose in indianapolis: Estimating trends using multilevel bayesian models. *Addictive behaviors* **86**, 4–10 (2018)
38. Ray, B., Quinet, K., Dickinson, T., Watson, D.P., Ballew, A.: Examining fatal opioid overdoses in marion county, indiana. *Journal of urban health* **94**(2), 301–310 (2017)
39. Ray, B.R., Lowder, E.M., Kivisto, A.J., Phalen, P., Gil, H.: Ems naloxone administration as non-fatal opioid overdose surveillance: 6-year outcomes in marion county, indiana. *Addiction* **113**(12), 2271–2279 (2018)
40. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover’s distance as a metric for image retrieval. *International journal of computer vision* **40**(2), 99–121 (2000)
41. Rudd, R.A., Paulozzi, L.J., Bauer, M.J., Burleson, R.W., Carlson, R.E., Dao, D., Davis, J.W., Dudek, J., Eichler, B.A., Fernandes, J.C., et al.: Increases in heroin overdose deaths 28 states, 2010 to 2012. *MMWR. Morbidity and mortality weekly report* **63**(39), 849 (2014)
42. Ruhm, C.J.: Corrected us opioid-involved drug poisoning deaths and mortality rates, 1999–2015. *Addiction* **113**(7), 1339–1344 (2018)
43. Seth, P., Scholl, L., Rudd, R.A., Bacon, S.: Overdose deaths involving opioids, cocaine, and psychostimulants united states, 2015–2016. *Morbidity and Mortality Weekly Report* **67**(12), 349 (2018)
44. Shirota, S., Gelfand, A.E., et al.: Space and circular time log gaussian cox processes with application to crime event data. *The Annals of Applied Statistics* **11**(2), 481–503 (2017)
45. Strickler, G.K., Zhang, K., Halpin, J.M., Bohnert, A.S., Baldwin, G., Kreiner, P.W.: Effects of mandatory prescription drug monitoring program (pdmp) use laws on prescriber registration and use and on risky prescribing. *Drug and Alcohol Dependence* (2019)
46. Von Luxburg, U.: A tutorial on spectral clustering. *Statistics and computing* **17**(4), 395–416 (2007)
47. Warner, M., Hedegaard, H.: Identifying opioid overdose deaths using vital statistics data (2018)
48. Wysowski, D.K.: Surveillance of prescription drug-related mortality using death certificate data. *Drug Safety* **30**(6), 533–540 (2007)
49. Xu, H., Zha, H.: A dirichlet mixture model of hawkes processes for event sequence clustering. In: *Advances in Neural Information Processing Systems*. pp. 1354–1363 (2017)
50. Yuan, B., Li, H., Bertozzi, A.L., Brantingham, P.J., Porter, M.A.: Multivariate spatiotemporal hawkes processes and network reconstruction. *arXiv preprint arXiv:1811.06321* (2018)
51. Zammit-Mangion, A., Dewar, M., Kadirkamanathan, V., Sanguinetti, G.: Point process modelling of the afghan war diary. *Proceedings of the National Academy of Sciences* **109**(31), 12414–12419 (2012)

52. Zhuang, J., Ogata, Y., Vere-Jones, D.: Analyzing earthquake clustering features by using stochastic reconstruction. *Journal of Geophysical Research: Solid Earth* **109**(B5) (2004)