Forecasting Heroin Overdose Occurrences from Crime Incidents

Ali Mert Ertugrul[1,2], Yu-Ru Lin[1], Christina Mair[1], and Tugba Taskaya Temizel[2]

[1] University of Pittsburgh, Pittsburgh PA, USA
{ertugrul, yurulin, cmair}@pitt.edu
[2] Middle East Technical University, Ankara, Turkey
ttemizel@metu.edu.tr

Abstract. Opioid overdoses continue to worsen in the United States, with rapid increases in overdose deaths involving heroin. This crisis, recognized as "opioid epidemic", has widespread consequences across every region and every demographic group. To enhance the overdose surveillance and to identify the areas in need of prevention effort, in this work, we explore the forecasting capability of heroin overdose occurrences using real-time crime data. Prior works suggested different types of links between the overdose occurrences and criminal activities, such as financial motives and common causes. Grounded on these observations, we present a model that utilizes the spatiotemporal structure of the crime incidents to forecast future heroin overdose occurrences. Results show that, our method achieves better performance, with significantly lower errors (in terms of RMSE and MAE) compared with the baseline method. Our method also allows for meaningful interpretation from both spatial and temporal aspects, including identifying predictive hotspots, local and global contributions, and informative features.

Keywords: Forecasting heroin overdoses · Spatiotemporal modeling · Modeling crime dynamics

1 Introduction

Opioid use disorders (OUD) and overdose rates in the United States have increased at an alarming rate since the past decade [22]. Overdose deaths involving prescription opioids have been continuously rising since the 1990s; heroin overdose deaths have sharply increased since 2010 [19, 4]. The age-adjusted rate for drug poisoning deaths involving heroin nearly quadrupled between 2000 and 2013 [10], and deaths from drug overdose are now the top cause of injury-related death in the United States [3]. The rate of growth of OUD and overdose, combined with the number of impacted individuals in the United States, has led many to classify this as an "opioid epidemic" [13].

Detailed assessments of OUD and overdose growth associated with population subgroups and spatial patterns of spread require consistently collected and geographically well-resolved space-time data [8, 7]. In the United States there is no systematic monitoring of drug abuse and dependence at either a regional or state level. The only existing estimates of incidence and prevalence rates are meant to represent the entire country (e.g., NSDUH) or are limited to a small number of communities [8]. One such database is City of Cincinnati opioid overdose dataset, which contains the daily opioid overdose incidents with location information. In this work, we utilize this dataset to explore the links between opioid overdoses and other social phenomena.

Highlighting the links between various social phenomena and opioid use has drawn significant attention. Among them, a number of studies have identified a relationship between the opioid use and crime. Bennett et al. stated that the people dependent on heroin or other types of opiates are disproportionately involved in criminal activities [1] in particular for crimes committed for financial gain [17]. Furthermore, Hammersley et al. [9] suggested that opportunities for drug use increase with involvement in criminal behavior. Seddon et al. [21] indicated that crime and drug use share common set of causes and they co-occur together due to this set of causes. Furthermore, routine activity theory requires *a suitable target*, *a likely offender* and *the absence of capable guardian* for crime to occur. Since *likely offenders* can reach *suitable targets* in different locations easily, the locations of drug-related arrests and opioid overdose incidents may not be in the same location but exhibit spatial lag effects. Given the relationship between the crime and opioid use, in this study, we aim to explore the forecasting capability of heroin overdose occurrences using real-time crime data. More specifically, we model the spatiotemporal patterns of the crime incidents to forecast future heroin overdose occurrences.

There have been machine learning techniques that employed temporal, spatial and spatiotemporal dependencies for event forecasting. Most of these works focused on predicting event occurrences instead of volume/count from digital traces. Among them, several studies utilized logistic regression (LR) to detect/forecast events using social media data relevant to crime [6], civil unrest [14], and anomalies [16]. Ramakrishnan et al. [18] proposed a framework to forecast civil unrest events from a variety of data sources employing LR with LASSO. Moreover, Ning et al. [15] suggested a protest forecasting approach from articles based on multiple instance learning. It jointly predicted events and identified event precursors. Zhao et al. proposed spatiotemporal event forecasting using modified Hidden Markov Model [23] and multi-task learning [24, 25]. Furthermore, Zhao et al. [26] presented an approach capable of distant-supervision of heterogeneous multi-task learning for multi-lingual spatial event forecasting. However, a great majority of these studies primarily considered forecasting performance instead of interpretation of the events over time and across space. Also, the interactions between spatial and temporal dimensions were mostly overlooked. In this work, we present a model that employs the spatiotemporal structure of the crime incidents to forecast future heroin overdose occurrences. Our method also allows for meaningful interpretation from both spatial

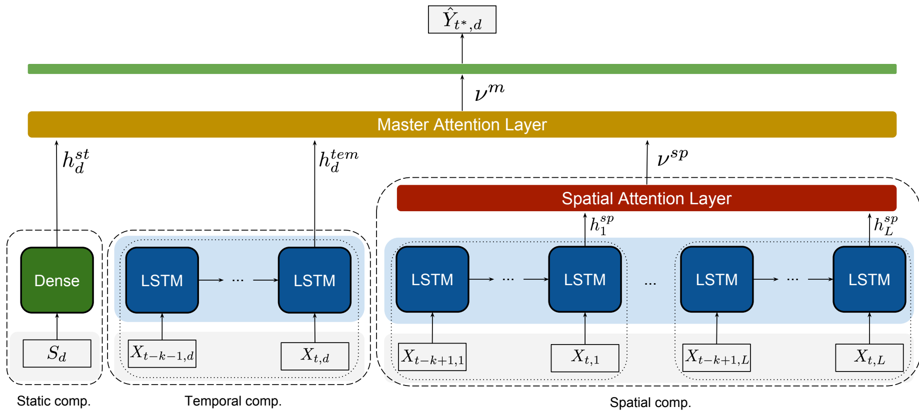


Fig. 1: Architecture. It mainly consists of three components namely static, temporal and spatial.

and temporal aspects, including identifying predictive hotspots, local and global contributions, and informative features.

2 Method

2.1 Problem Definition

Suppose that there exist L locations (e.g., neighborhoods, cities) and each location l can be represented by a combination of dynamic and static features. Let $\mathcal{X}_{t-k+1:t} = \{X_{t-k+1:t,l}, l = 1, 2, \ldots, L\}$ denotes the historical collection of dynamic features from all locations within a *time window* with size k up to time t, where $X_{t,l}$ is set of the aggregated dynamic features (e.g. number of crime incidents) at time t (e.g. week, month) and location l. Furthermore, let $S = \{S_l, l = 1, 2, \ldots, L\}$ be the collection of the static features (e.g. population of a location, median household income) which either remain the same or change is visible in a long period of time. Moreover, $Y_{t^*,l} \in \mathbb{N}$ denotes the number of heroin overdoses occurred at a future time t^* and location l.

Our purpose is to forecast $Y_{t^*,l}$ – the number of heroin overdose occurrences at the future time t^* and location l where the time difference between t^* and t is the lead time for forecasting – from the static and dynamic features (historical crime activities) obtained from location l with the dynamic features from all other locations. As a result, the forecasting problem can be formulated as learning the mapping function $f(S_d, \mathcal{X}_{t-k+1:t}) \rightarrow Y_{t^*,d}$ from the static and dynamic features, to the number of heroin overdose occurrences at the time t^* for a target location d.

2.2 Model

Our model consists of three main components namely static component, temporal component and spatial component, as shown in Fig 1. Static component encodes the static features of the target location d with a fully connected layer. The output of static component is h_d^{st}. On the other hand, temporal and spatial components are designed to capture local and global crime dynamics using Long Short Term Memory (LSTM) [11] as building blocks. While the former models the local dynamic features, the latter models the spatiotemporal contribution of dynamic features of all locations. For the temporal component, we use a single LSTM network. On the other hand, the spatial component includes separate LSTM networks for each location to model its temporal dynamics. The LSTM outputs inside temporal and spatial components are h_d^{tem} and $\{h_1^{sp}, h_2^{sp}, \ldots, h_L^{sp}\}$, respectively.

We extend the two-level attention mechanism first introduced in [5]. The key idea is to identify predictive hotspots and differentiate the contribution of static and dynamic features on forecasting heroin overdoses. The first attention layer, namely spatial attention layer, is located in the spatial component and on top of $\{h_1^{sp}, h_2^{sp}, \ldots, h_L^{sp}\}$. Its purpose is to highlight the locations that have more contribution on forecasting heroin overdoses in the target location. Therefore, this attention layer identifies the predictive hotspots. ν^{sp} is the output of spatial attention layer which summarizes the information from of all locations. Second, we present master attention layer on top of the outputs of three components, which are h_d^{st}, h_d^{tem} and ν^{sp}. The purpose of master attention is to differentiate the contributions of static, local dynamic and global dynamic features. For instance, while static features are more predictive in some locations, the local or global crime dynamics are more informative in another regions. Therefore, this attention adjusts and summarizes the contributions of different components in its output ν^m. Then, a hidden layer with ReLU activation function is applied to forecast the number of heroin overdose occurrences as $\hat{Y}_{t^*,d}$ in the target location d at the future time t^*.

For the training of the network, we use the mean squared error as the loss function. We also employ Group Lasso regularization to select informative features inspired from [20]. Each input neuron in the all components of the model is considered as a separate group.

2.3 Features

We utilize two types of features namely static features and dynamic features.
Static features are obtained from census data of City of Cincinnati and they include demographics and economic status about the neighborhoods. These features change slowly over time. Among the demographics features, for each neighborhood we extract population, distribution of genders and distribution of races (White alone, Black or African American alone, American Indian and Alaska Native alone, Asian alone, Native Hawaiian and Other Pacific Islander alone, Two or More Races). Furthermore, we employ median household income, per capita income and poverty as the static features showing the economic status of the neighborhoods. As a result, we obtain a total of 12 static features. Note that, we normalize the gender, race related features and poverty by dividing them to

total population of corresponding neighborhoods. We also apply z-score normalization for median household income and per capita income, and log transform for population before training our architecture.

Dynamic features are to capture the crime dynamics of the City of Cincinnati that may be predictive for heroin overdose occurrences. We obtain dynamic features from crime incidents dataset. It includes crime incidents with a unique crime incident number. Each crime incident may contain one or more Uniform Crime Reporting (UCR) types (high-level offense). The available UCR types in the dataset are Part 2 Minor, Theft, Burglary/Breaking Entering, Robbery, Aggravated Assaults, Rape and Unauthorized Use. For each neighborhood and each time unit, we aggregate total number of unique incidents, total number of incidents of all UCR types and total number of incidents for each of eight individual UCR types aforementioned above. As a result, we obtain a total of 9 dynamic features for each neighborhood for a given time unit. Note that, for a given neighborhood we first normalize total number of incidents of each individual UCR type by dividing it to the total number of incidents of all UCR types. Then, we apply z-score normalization to total number of unique incidents and total number of incidents of all UCR types before training the architecture.

3 Experiments

3.1 Dataset

We employ three datasets namely Heroin Overdoses[1], Police Data Initiative (PDI) Crime Incidents[2] and 2010 US Census Data for City of Cincinnati[3]. The first one is used to form ground-truth labels (heroin overdose occurrences) while the second one is processed to generate dynamic predictors. The incidents between Aug 2015 and May 2018 (included) are considered in these two datasets. The retrieval date for both datasets is June 1st, 2018. Moreover, census data of the City of Cincinnati is also considered as the static predictors which change slowly over time and may provide significant information on forecasting future heroin overdose occurrences.

3.2 Settings

We used ‘month’ as the time unit and ‘neighborhood’ in City of Cincinnati as the location unit. The data (both heroin overdose and crime incidents) spanning the first two years was used as the training set. The data of next 3 months was used for validation and the last 5 months was used as test set. The hidden unit size for LSTM units in the model is 16. The window size is set to be $\{1, 2, 3\}$ and the lead time is set to be $\{1, 2\}$. Also, the model is trained with mini-batch stochastic gradient descent (SGD).

[1] https://insights.cincinnati-oh.gov/stories/s/Heroin/dm3s-ep3u/

[2] https://data.cincinnati-oh.gov/Safer-Streets/PDI-Police-Data-Initiative-Crime-Incidents/k59e-2pvf

[3] https://www.cincinnati-oh.gov/planning/reports-data/census-demographics/

Table 1: Forecasting results

	RMSE	MAE	Pearson	Spearman's Rank
Baseline	2.954	2.003	0.701*	0.691*
Our model	**2.672**	**1.473**	**0.715***	**0.722***

* $p < 0.001$

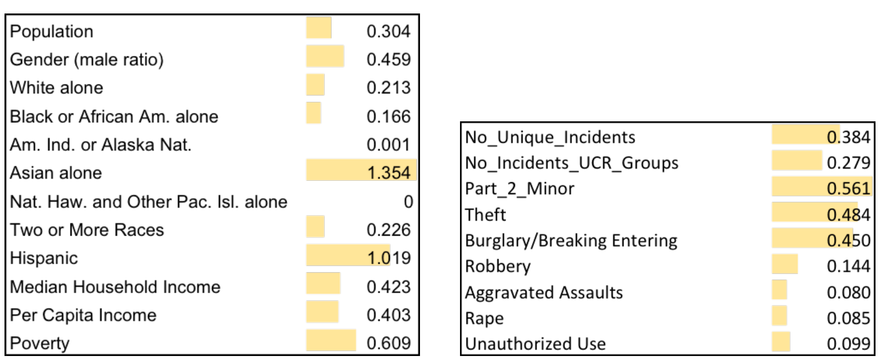
4 Results

4.1 Performance Comparison

To evaluate our model, we compare it with a baseline linear regression model in which static features and dynamic features are concatenated to predict the number of heroin overdose occurrences. For the quantitative evaluation, we use root mean squared error (RMSE), mean absolute error (MAE), Pearson and Spearmans rank correlation. As given in Table 1, our model achieves lower RMSE (2.672) and lower MAE (1.473) compared to baseline (2.954 and 2.003). Furthermore, the values of Pearson and Spearmans rank correlation are 0.715 & 0.722 for our method and 0.701 & 0.691 for the baseline, respectively. The correlation results for our method are also superior to the results of the baseline. Note that, correlation values for both models are statistically significant.

4.2 Analysis of Features

Since we incorporate Group Lasso regularization into our model, we expect our model to select informative features for forecasting heroin overdoses. Within the scope of this study, we analyze the static feature weights and local dynamic feature weights to see which features are more informative on forecasting heroin overdose occurrences. Fig. 2 shows the mean absolute values of corresponding feature weights. According to the results, Asian and Hispanic population are found as important demographic predictors on forecasting future heroin overdoses in a



(a) Static feature weights (b) Local dynamic feature weights

Fig. 2: Importance of static and local dynamic features

neighborhood. Furthermore, Boardman et al. [2] states that a lack of economic resources may create communities with greater vulnerability to substance use. Moreover, communities with a higher concentration of economic stressors (e.g., low median income) may be particularly vulnerable to abuse of opioids as a way to manage chronic stress and anxiety and mood disorders [12]. Our findings are consistent with the literature and our model identifies poverty and median household income as important economical features as shown in Fig. 2.

Theories of the drugs - crime connection predict that certain kinds of offenses, such as shoplifting, theft, robbery, burglary and prostitution are more likely than others to be associated with drug use and they might be committed to raise funds to purchase drugs [1]. Consistent to this statement, we observe that the proportion of the total number of certain crime groups, namely 'Part 2 Minor', 'Theft' and 'Burglary/Breaking Entering', are found to be more informative by our model compared to the other crime groups such as 'Aggravated Assaults' and 'Rape'. The total number of unique crime incidents and incidents of all UCR types in a neighborhood are also crucial for the prediction.

4.3 Identification of Predictive Hotspots

Our model provides interpretability while forecasting the heroin overdose occurrences in a specific neighborhood in terms of two aspects. First, our model allows us to discover the predictive hotspot neighborhoods. Second, it enables us to examine the contribution of local and global crime dynamics, and static features on forecasting future heroin overdose occurrences. To analyze the predictive hotspots, we examine the weights of spatial attention and plot these hotspots in Fig. 3 based on their attention weights. According to the figure, there are a few neighborhoods whose crime dynamics have global contribution on forecasting heroin overdoses in other neighborhoods. While *East Price Hill* (13) and *Westwood* (49) are the globally most predictive ones, *West Price Hill* (48) and *Walnut Hills* (46) have little contribution on forecasting.

Furthermore, to analyze contribution of local and global crime dynamics, and static features on forecasting overdoses, we analyze the weights of the master attention on a randomly selected neighborhood namely *Downtown* (11) in a specific time unit (May 2018). We observe that, while the contribution of local and global crime dynamics are 0.449 and 0.226, respectively, and static features contribute 0.325. In other words, the most predictive information stems from the local crime dynamics of *Downtown* itself. On the other hand, predictive hotspots has less contribution.

In addition to predictive hotspots analysis on whole data, we further decompose the heroin overdoses and crime incidents datasets into two equal parts, spanning the time intervals (Aug 2015 - Dec 2016) and (Jan 2017 - May 2018), to observe whether there exists a change in predictive hotspots based on the time interval. We train separate models (85% training set and 15% test set) for each time interval using our architecture. Fig. 4 indicates the predictive hotspots discovered with the analysis of spatial attention weights of two models. We observe that East Price Hill (13), Westwood (49) and West Price Hill (48) are

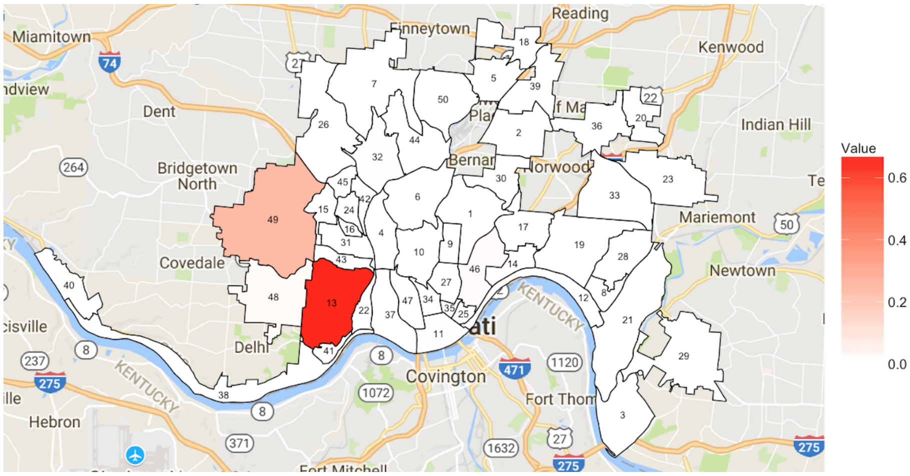


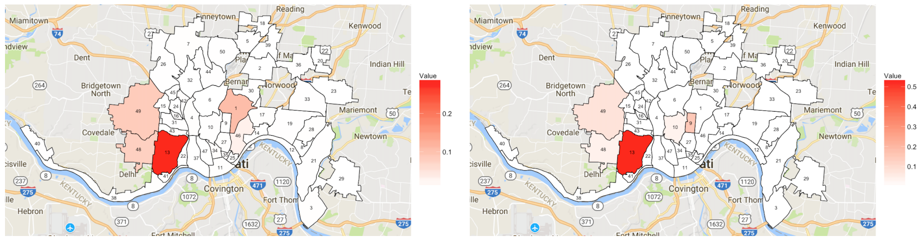
Fig. 3: Predictive hotspot neighborhoods. The neighborhoods with darker red color have more global contribution on forecasting heroin overdose occurrences on the other neighborhoods. East Price Hill (13) and Westwood (49) are the globally most contributing neighborhoods. West Price Hill (48) and Walnut Hills (46) have also little contribution on forecasting.

the common predictive hotspots for both time intervals. They are the neighborhoods where the number of crime incidents are the highest among all the neighborhoods. Remember that they are also predictive hotspots discovered by the model trained using whole dataset. Furthermore, Avondale (1) and Walnut Hills (46) seem predictive hotspots before 2017 and they disappear after 2017. When we compare the number of crime incidents in both neighborhoods before and after 2017, we see a decrease in the number of incidents after 2017. Moreover, Corryville (9) has a considerable contribution to forecast heroin overdoses globally after 2017. There is a significant increase in the number of crime incidents in this neighborhood after 2017. Accordingly, we can infer that immediate changes in the number of crime incidents in a neighborhood can be an indicator for being a predictive hotspot.

5 Discussion and Future work

In this work, we present a model that employs the spatiotemporal structure of the crime incidents to forecast future heroin overdose occurrences. Results indicate that, our method achieves better performance, with significantly lower errors (in terms of RMSE and MAE) compared with the baseline method. Our method also provides interpretation capability from both spatial and temporal aspects, including identifying predictive hotspots, local and global contributions, and informative features.

Furthermore, there exist several limitations in our work. First, our architecture captures the spatial relationships (crime dynamics) among the neigh-



(a) Aug 2015 - Dec 2016 (b) Jan 2017 - May 2018

Fig. 4: Change in predictive hotspots. The neighborhoods with darker red color have more global contribution on forecasting heroin overdose occurrences on the other neighborhoods for given time interval. The common predictive hotpots for both time intervals are East Price Hill (13), Westwood (49) and West Price Hill (48). While Avondale (1) and Walnut Hills (46) are predictive hotspots before 2017, Corryville (9) and CUF (10) contribute globally after 2017.

borhoods as a whole. It does not model pairwise relationship between target neighborhood and any other. Second, we utilize static features extracted from the US 2010 census data. Although static features are expected to change slowly over time, a more up-to-date census data may yield more accurate results. As a future work, we plan to improve our architecture in a way that it considers spatial relationships between all pairs of locations. We also plan to utilize static features extracted from more recent sources.

References

1. Bennett, T., Holloway, K., Farrington, D.: The statistical association between drug misuse and crime: A meta-analysis. Aggression and Violent Behavior **13**(2), 107–118 (2008)
2. Boardman, J.D., Finch, B.K., Ellison, C.G., Williams, D.R., Jackson, J.S.: Neighborhood disadvantage, stress, and drug use among adults. Journal of health and social behavior pp. 151–165 (2001)
3. Centers for Disease Control and Prevention: Web-based Injury Statistics Query and Reporting System (WISQARS). http://www.cdc.gov/injury/wisqars/fatal.html, accessed: 2018-05-01
4. Compton, W.M., Jones, C.M., Baldwin, G.T.: Relationship between nonmedical prescription-opioid use and heroin use. New England Journal of Medicine **374**(2), 154–163 (2016)
5. Ertugrul, A.M., Lin, Y.R., Yan, M., Chung, W.T., Li, A.: Activism via attention: Interpretable spatiotemporal learning to forecast protest activities (2018)
6. Gerber, M.S.: Predicting crime using twitter and kernel density estimation. Decision Support Systems **61**, 115–125 (2014)
7. Gruenewald, P.J.: Geospatial analyses of alcohol and drug problems: empirical needs and theoretical foundations. GeoJournal **78**(3), 443–450 (2013)
8. Gruenewald, P.J., Ponicki, W.R., Remer, L.G., Waller, L.A., Zhu, L., Gorman, D.M.: Mapping the spread of methamphetamine abuse in california from 1995 to 2008. American journal of public health **103**(7), 1262–1270 (2013)

9. Hammersley, R., Forsyth, A., Morrison, V., Davies, J.B.: The relationship between crime and opioid use. Addiction **84**(9), 1029–1043 (1989)
10. Hedegaard, H., Chen, L.H., Warner, M.: Drug-poisoning deaths involving heroin: United states, 2000-2013 (2015)
11. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997)
12. King, N.B., Fraser, V., Boikos, C., Richardson, R., Harper, S.: Determinants of increased opioid-related mortality in the united states and canada, 1990–2013: a systematic review. American journal of public health **104**(8), e32–e42 (2014)
13. Kolodny, A., Courtwright, D.T., Hwang, C.S., Kreiner, P., Eadie, J.L., Clark, T.W., Alexander, G.C.: The prescription opioid and heroin crisis: a public health approach to an epidemic of addiction. Annual review of public health **36**, 559–574 (2015)
14. Korolov, R., Lu, D., Wang, J., Zhou, G., Bonial, C., Voss, C., Kaplan, L., Wallace, W., Han, J., Ji, H.: On predicting social unrest using social media. In: ASONAM. pp. 89–95 (2016)
15. Ning, Y., Muthiah, S., Rangwala, H., Ramakrishnan, N.: Modeling precursors for event forecasting via nested multi-instance learning. In: Proceedings of the 22nd ACM SIGKDD. pp. 1095–1104 (2016)
16. Panagiotou, N., Zygouras, N., Katakis, I., Gunopulos, D., Zacheilas, N., Boutsis, I., Kalogeraki, V., Lynch, S., O'Brien, B.: Intelligent urban data monitoring for smart cities. In: ECML-PKDD. pp. 177–192 (2016)
17. Pierce, M., Hayhurst, K., Bird, S.M., Hickman, M., Seddon, T., Dunn, G., Millar, T.: Quantifying crime associated with drug use among a large cohort of sanctioned offenders in england and wales. Drug & Alcohol Dependence **155**, 52–59 (2015)
18. Ramakrishnan, N., Butler, P., Muthiah, S., Self, N., Khandpur, R., Saraf, P., Wang, W., Cadena, J., Vullikanti, A., Korkmaz, G., et al.: 'beating the news' with embers: forecasting civil unrest using open source indicators. In: Proceedings of the 20th ACM SIGKDD. pp. 1799–1808 (2014)
19. Rudd, R.A., Aleshire, N., Zibbell, J.E., Matthew Gladden, R.: Increases in drug and opioid overdose deathsunited states, 2000–2014. American Journal of Transplantation **16**(4), 1323–1327 (2016)
20. Scardapane, S., Comminiello, D., Hussain, A., Uncini, A.: Group sparse regularization for deep neural networks. Neurocomputing **241**, 81–89 (2017)
21. Seddon, T.: Drugs, crime and social exclusion: social context and social theory in british drugs–crime research. British Journal of Criminology **46**(4), 680–703 (2005)
22. Warner, M., Chen, L.H., Makuc, D.M., Anderson, R.N., Miniño, A.M.: Drug poisoning deaths in the united states, 1980-2008. NCHS data brief (81), 1–8 (2011)
23. Zhao, L., Chen, F., Lu, C.T., Ramakrishnan, N.: Spatiotemporal event forecasting in social media. In: SDM. pp. 963–971 (2015)
24. Zhao, L., Sun, Q., Ye, J., Chen, F., Lu, C.T., Ramakrishnan, N.: Multi-task learning for spatio-temporal event forecasting. In: Proceedings of the 21th ACM SIGKDD. pp. 1503–1512 (2015)
25. Zhao, L., Wang, J., Chen, F., Lu, C.T., Ramakrishnan, N.: Spatial event forecasting in social media with geographically hierarchical regularization. Proceedings of the IEEE **105**(10), 1953–1970 (2017)
26. Zhao, L., Wang, J., Guo, X.: Distant-supervision of heterogeneous multitask learning for social event forecasting with multilingual indicators. In: AAAI (2018)