
Deep Learning & the Opioid Epidemic: Estimating Opioid-Related Mortality Risk in US Counties with Twitter Data

Zoe-Alanah M Robert

Department of Computer Science
Stanford University
zrobert7@alumni.stanford.edu

Abstract

The widespread misuse and addiction to opioids - which include prescription pain relievers, heroin, and synthetic opioids like Fentanyl - has become a national crisis in the US. This ‘Opioid Epidemic’ has far reaching side effects, in everything from public health to economic and social well being. A number of government and private institutions have publically available historical data which, given the right data analysis tools, can help us understand which areas of the nation have the highest risk of a local overdose epidemic. As many opioid addicts ultimately turn to illegal forms of the drug, there is a known marketplace on social media. Additionally, pop culture has effectively popularized the use and abuse of opioids. These, along with existing research, suggest that Twitter data might hold some predictive power in assessing community risk of an opioid overdose outbreak. This project analyzes the use of Twitter data and historical public health data with neural networks to predict future community risk. Overall, historical public health data proved more useful and better at generalizing to unseen counties (achieving a mean squared error (MSE) of 0.01 and R^2 of 0.57 on test data). However, the Twitter data showed some correlation with future community risk, suggesting the the potential for further research.

1 Introduction

The opioid epidemic is plaguing the US - on average, over 100 Americans die from an opioid-related overdose every day^[1]. Many addictions form after patients are prescribed a legal opioid by a doctor and form a dependency. After the prescription runs out, patients are often unable to pay for the same quality drug and turn to cheaper opioids sold illegally. If community risk (likelihood of future deaths by opioid overdose) can be accurately predicted, at-risk communities can be flagged early on so that government prevention resources can be targeted towards these communities. Additionally, according to the National Institute on Drug Abuse, an opioid overdose (from opioids that include heroin, Oxycontin, Fentanyl, Morphine, Vicodin, Percocet and Methadone) can be effectively reversed if the drug Narcan (naloxone) is given soon after overdose - while the patient is still breathing^[2]. A predictive model which suggests which counties in the US might be at high risk for an overdose outbreak could help re-distribute Narcan, which is often in short supply, to those who need it most. Thus, I have built a deep model for identifying opioid user groups and at-risk communities. More specifically, I focused on two learning tasks: Historical and Twitter.

For the Historical task, the input is a 98-feature vector containing opioid prescribing rates from 2008 to 2016, the 2015 population estimate for that county, and raw death counts per COD (Cause of Death

code) for the years 2008 to 2015. The output is the adjusted 2016 Opioid-related mortality rate per 10000 persons in the same county. This mortality rate acts as a risk level for the county as it suggests either frequent overdoses or lack of Narcan availability to patients. For the more experimental Twitter Task, the input is a word2vec embedding of a collection of tweets posted in that county in a single week which contain opioid-relevant terms. These terms include both medical and “street” names for common opioids. Some examples include: ‘Fentanyl’, ‘Codiene’, ‘Percocet’, ‘percs’, ‘Chinawhite’, etc. The output is the same as the Historical Task: the adjusted 2016 Opioid-related mortality rate. Although this project involves two distinct regression prediction tasks, similarly structured neural networks proved effective for both. I experimented with different architectures, loss functions, and generalization techniques. Ultimately, the best models for the Historical task show potential to directly predict opioid risk at the county level within some threshold, while the same prediction task directly from Twitter data proved much more difficult. I hypothesize that the combination of data inputs along with additional tweet filtering would produce even better results.

2 Related work

2.1 Dataset scale

A number of very recent publications have attempted to analyze social media data and medical records directly related to the Opioid Crisis. Che, et al. published a study in 2017 in which they were able to successfully classify opioid users and to “extract key factors for different opioid user groups” using RNNs on a dataset of over one hundred thousand opioid users^[9]. Similarly, Gebert, et al. in 2018 looked at hand-selected features from autopsy data and fake medical records for a single county in Pennsylvania and used both linear models and neural networks to predict (with high accuracy) whether the COD was opioid overdose or not^[10]. In both of these publications, researchers successfully applied deep learning to better understand trends in the Opioid Epidemic. While the inspiration for my model architectures did come from the neural network in Gerbert, et al, it is important to note that these papers were working at the individual scale while I am looking at the county scale.

2.2 Using Tweets

There has also been a plethora of research tying social media data to trends in public health. In 2017, Google showed that user search trends “can detect early signs of diabetes by monitoring combinations of keywords” in Google search histories^[11]. This hypothesis that user data acts a public health signal extends to the Opioid Crisis. In 2017, Mackey et al. were able to automatically identify nearly two thousand tweets marketing the sale of controlled substances online^[12]. In 2016, Sarker et al. identified a process to automatically classify tweets as marketing the illegal abuse of prescription medication, finding that “clear signals of medication abuse can be drawn from Twitter posts”^[13]. A 2015 report by Chan et al. was able to use Twitter data to identify public sentiment regarding the Opioid Crisis and behavior patterns from tweets^[14]. A 2018 study by Graves, et al. arrives at perhaps the most powerful finding yet: “Regional differences in opioid-related topics reflect geographic variation in the content of Twitter discussion about opioids. Analysis of Twitter data also produced topics significantly correlated with opioid overdose death rates”^[15]. Based on the success of these reports and my own intuition on the role of social media in the Opioid Crisis, I chose to look into opioid-related Twitter data as a means of directly predicting future opioid-related mortality in the US on the county level. Noticing the lack of existing works which utilize deep learning, I chose to distinguish my work by applying neural networks to the problem of opioid-related tweets and to focus on prediction at the county level.

3 Dataset and Features

There is no pre-existing, clean dataset related to county-level opioid crisis risk assessment so the dataset for this project was curated from several sources, outlined below:

A. Historical Opioid Prescribing Rates from the Center for Disease Control (CDC)^[3]: This contains the opioid prescribing rate per 100 people (keep in mind some patients are prescribed more than one opioid at a time) per county in the US, per year, for the years 2008 to 2016.

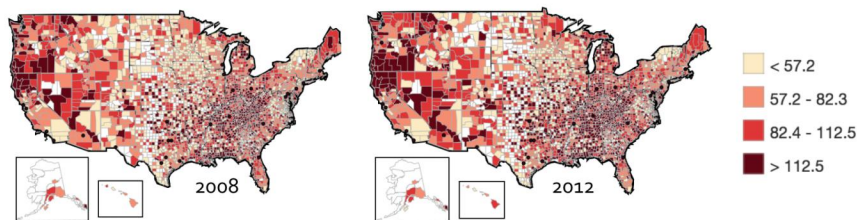


Figure 1: Opioid prescriptions per 100 people by US county^[3]

'Them perc's put me in my feelings every time 🙄🙄'

'Taskforce Completes Successful Opioid Bust Near Tribal Reservations in Arizona - Native News Online <link>'

'@kasie When I was 1st prescribed Oxy for pain, I was told it was not addictive. For the next 6 years I was hopeless...'

'where da perc's at'

'Order quality Oxycodone, Percocet, Dilaudid, Codeine, xanax, viagra, Nembutal, Ritalin, Valium, Roxicodone and many... <link>'

Figure 2: Example opioid-related tweets

B. Opioid-Related Deaths from CDC Wonder^[4]: This table, requested from CDC Wonder, contains the crude and age-adjusted mortality rates per county, per year for the years 2008 to 2016. The data is further split into the Cause of Death Codes (CODs) related to Opioids as outlined by drugabuse.gov^[5]

C. Tweets Mentioning Opioids from Twitter API^[6]: a database of tweets scraped from the Standard Search API that mention any of the official “street” or medical names for Opioids as outlined by the Foundation for a Drug-Free World.^[7]

The data is collected and then distinguished on county: an (x, y) pair represents a single county. For the Historical task, x_i is a feature vector containing county i 's 2015 population, death counts for each of the 11 opioid-related death codes for each of the 8 years, and a prescribing rate per 100 people for each of the 9 years. There are several counties / years missing prescribing rates. In preprocessing, these NaNs are converted to zero values. X is then normalized before being split into train/dev/test sets of sizes 461/86/61.

For the Twitter task, the free Standard Search API is used to obtain any tweets with a (latitude, longitude) tag within a 20 mile radius (radius corresponding to the average US county area: 1208 mi²) of the official county center, as outlined by the US census and official FIPS county codes^[8]. Unfortunately, due to the limitations of the Standard Search API, only tweets from the past seven days are accessible. Thus, in this project 2018 tweets are used to predict 2016 opioid-related mortality rates. However, given sufficient financial resources, a full year of tweets and the correct year (2015) could be used. After scraping all available, relevant tweets, I remove any links, remove non alphanumeric characters, and lowercase all characters. As some counties contain fewer relevant tweets than others, all county-level embeddings are padded to the max length, resulting in input vectors of 140300 features. The embeddings are produced from Google's pretrained word2vec model, available through gensim^[18]. Then, I normalize these 608 140300-feature word embeddings and use them as X input to the model.

For both learning tasks, y_i is a single float that represents the sum over all 11 raw opioid-related death code counts, adjusted to a mortality rate per 10000 persons for the year 2016. Only counties which have full 2016 mortality data are considered. There are currently 3144 counties in the US, but only 608 counties with sufficient y label data. Unfortunately, data augmentation is difficult given the realistic geographic nature of the data set and the lack of clarity regarding its underlying distribution.

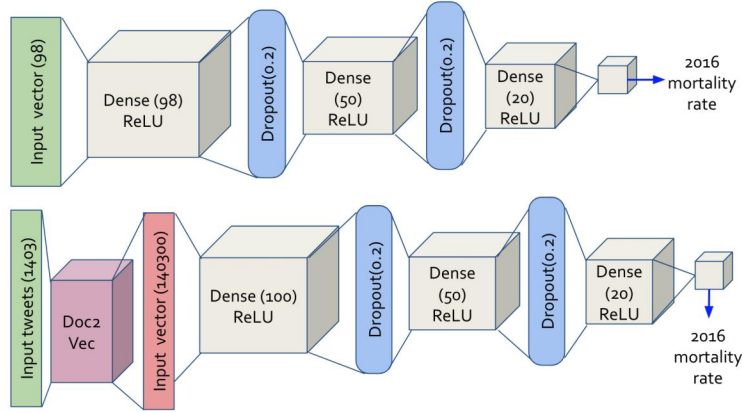


Figure 3: *NN4 model architecture (top) and NN3_twitter_dropout model architecture (bottom) .*

Mean squared error $MSE = \frac{1}{n} \sum_{t=1}^n e_t^2$ Mean absolute error $MAE = \frac{1}{n} \sum_{t=1}^n |e_t|$

Figure 4: *Mathematical formulas for different loss functions.*

4 Methods

All models were implemented using Keras^[17], with Tensorflow as the backed. After implementing linear baseline models, the approach for solving this problem was to incrementally build up from small neural networks, adding a layer at a time to see the improvements on success metrics. The notation used for the different neural network model architectures experimented with is as follows: Each of NN_x (NN_1 , NN_2 , and NN_3) contain x dense layers with ReLU activation. The NN_4 structure is three sets of a dense layer with ReLU activation and a dropout layer. (See Figure 3). In addition to experimenting with dense layers and dropout, I compared models trained with different loss functions. The two loss functions used were mean squared error (MSE) and mean absolute error (MAE), the formulas for which are defined in Figure 4. Each of NN_x_{loss} uses the specified *loss* function. And $NN1_{twitter_{loss}dropout}$ is similar to $NN1_{twitter_{loss}}$ but with additional dropout layers after the dense layers as shown in Figure 3.

5 Results

Although there are over 3000 counties in the US, due to limited data availability, this set was reduced to 608 after preprocessing. Given the small data set and the relatively small input feature vector size for both tasks, a batch size of the full length of X was chosen. All training, dev, and test set input data was normalized. In all models, gradient clipping proved effective in reducing the effects of gradient

Historical Task			Twitter Task		
Model Architecture	Train R^2	Test R^2	Model Architecture	Train R^2	Test R^2
Baseline linear	0.37	0.27	Baseline linear	0.53	-0.58
NN1_mse	0.44	0.10	NN1_twitter_mse	-0.06	0.24
NN3_mse	0.65	0.38	NN3_twitter_mse	0.62	-0.40
NN4_mse	0.57	0.55	NN1_twitter_mae	0.48	-0.51
NN1_mae	0.59	0.50	NN3_twitter_mae	0.69	-0.38
NN3_mae	0.64	0.57	NN3_twitter_mse_dropout	0.61	0.12
NN4_mae	0.20	0.37	NN3_twitter_mae_dropout	0.60	0.02

Figure 5: *Results averaged over 10-fold cross-validation. For the Baseline linear, four separate linear models were run (linear regression, ElasticNet, Lasso, Ridge) and the results from the best model are displayed here.*

“cliffs.” Hyperparameter tuning resulted in a top learning rate of 0.001 for both learning tasks. In both models, overfitting to the training data was a big problem, although mitigated with the addition of dropout layers. For the Twitter task model, an added hyperparameter was the minimum count a word had to appear to be considered in the word2vec model. Ultimately, due to the smaller size of the dataset and unique nature of the opioid terms, a min count of 1 was selected. Adam optimization was chosen for both due to its learning speed. As this is a regression model, the primary metrics for this project were R^2 , mean squared error (MSE), mean absolute error (MAE), and mean absolute percent error (MAPE). For both learning tasks, combinations of these metrics as loss functions were explored - primarily focusing on MSE and MAE as shown in Figure 5. However, R^2 was heavily used to compare different architectures. Since the opioid mortality rate per 10000 persons is still often a decimal, small MSE and MAE values were not particularly useful as final architecture comparison metrics.

The NNx_mse results in the linear baseline models for the Historical Task (see Figure 5) suggest that increasing the number of dense layers improves training results, but that the model does not generalize well. With the addition of the Dropout layers in NN4_mse, we see a big improvement in the model’s performance on test data. Similarly in the models using MAE loss, we see improvement with the addition of fully connected layers. MAE loss over MSE loss leads to overall higher performance on this learning task. The relatively good performance of the best baseline linear model suggests that more than a few fully connected layers would lead to over-fitting.

Comparing the results from different model architectures, we see that the Twitter task is much more difficult to learn. The addition of dense layers between NN1_twitter_mse and NN3_twitter_mse improves training metrics but worsens test metrics, suggesting the need to reduce model complexity or improve the model’s ability to generalize. The use of MSE loss over MAE loss seems to improve performance. Finally, with many dense layers and the addition of dropout layers, we achieve the best test set performance in NN3_twitter_mse_dropout with an R^2 of 0.12. While this resulting test set performance is still relatively poor, the Twitter learning task itself is experimental and any correlation between tweets and future mortality rates in a county is significant.

6 Conclusion/Future Work

In conclusion, we see that future opioid-related mortality rate prediction from historical public health data at the county level is indeed possible with deep learning. Multiple fully-connected layers along with dropout layers help the model generalize to unseen counties. Too many epochs or too deep of a network reduce performance. Also, for the Historical learning task we see that mean absolute error is more effective than mean squared error. Additionally, we learned in this project that while Twitter data is not the most effective in predicting future mortality rates related to opioids, there is in fact a correlation between the Tweets in a county that mention opioid-related terms and the overdose rate in that county, as supported by previous works.

Given more time and resources, further improvements could be made in the Historical Task with the addition of various publicly available datasets. These include, but are not limited to: drug addiction facility admissions, poverty rates, median income, historical opioid price data, self reported use and misuse, etc. I would explore these along with different means of representing missing data elements other than zero initializations. For the Twitter task, a major hurdle in this project has been the lack of financial resources to access the Premium Twitter API - the Twitter data collected for this project was limited to the free Standard Search API, which only contains tweets from the last seven days. Ideally, I would use tweets from 2015 to predict 2016, or perhaps even a collection of tweets over the course of many years. Additionally, I believe it would be useful to explore tweet classification beyond the mention of an opioid related term. How can we filter out news articles about the Opioid crisis, song lyrics, and other less-relevant tweets? Effective classification would allow more granular tweet filtering. Using the current data set, I would like to try using a combination of historical and Twitter data as model input. Lastly, I would like to explore different loss functions. While mean squared error and mean absolute error are useful, depending on the use-case of the model, a rank-based loss function might be more effective.

7 Contributions

I worked individually for this project.

8 Code Appendix

The code for this project (excluding datasets) can be found on Github:

https://github.com/zrobert7/cs230_opioids

References

- [1] "Opioid Overdose," Centers for Disease Control and Prevention, 30-Aug-2017. [Online]. Available: <https://www.cdc.gov/drugoverdose/epidemic/index.html>. [Accessed: 07-Jun-2018].
- [2] National Institute on Drug Abuse, "Opioid Overdose Reversal with Naloxone (Narcan, Evzio)," NIDA, 04-Apr-2018. [Online]. Available: <https://www.drugabuse.gov/related-topics/opioid-overdose-reversal-naloxone-narcan-evzio>. [Accessed: 07-Jun-2018].
- [3] "Opioid Overdose," Centers for Disease Control and Prevention, 31-Jul-2017. [Online]. Available: <https://www.cdc.gov/drugoverdose/maps/rxcounty2008.html>. [Accessed: 07-Jun-2018].
- [4] "CDC WONDER", Wonder.cdc.gov, 2018. [Online]. Available: <https://wonder.cdc.gov/>. [Accessed: 07- Jun-2018].
- [5] "Opioid Summaries by State", Drugabuse.gov, 2018. [Online]. Available: <https://www.drugabuse.gov/drugs-abuse/opioids/opioid-summaries-by-state>. [Accessed: 07- Jun- 2018].
- [6] "Premium search", Developer.twitter.com, 2018. [Online]. Available: <https://developer.twitter.com/en/docs/tweets/search/overview/premium>. [Accessed: 07- Jun- 2018].
- [7] Drugfreeworld.org, 2018. [Online]. Available: <https://www.drugfreeworld.org/drugfacts/prescription/opioids-and-morphine-derivatives.html>. [Accessed: 07- Jun- 2018].
- [8] G. Branch, "2010 FIPS Code Files for Counties - Geography - U.S. Census Bureau", Census.gov, 2018. [Online]. Available: <https://www.census.gov/geo/reference/codes/cou.html>. [Accessed: 07- Jun- 2018].
- [9] Che, Zhengping, et al. "Deep Learning Solutions for Classifying Patients on Opioid Use." AMIA Annual Symposium Proceedings. Vol. 2017. American Medical Informatics Association, 2017.
- [10] Gebert, Theresa, Shuli Jiang, and Jiaxian Sheng. "Characterizing Allegheny County Opioid Overdoses with an Interactive Data Explorer and Synthetic Prediction Tool." arXiv preprint arXiv:1804.08830 (2018).
- [11] Tkachenko, Nataliya, et al. "Google Trends can improve surveillance of Type 2 diabetes." Scientific reports 7.1 (2017): 4993.
- [12] Mackey, Tim K., et al. "Twitter-Based Detection of Illegal Online Sale of Prescription Opioid." American journal of public health 107.12 (2017): 1910-1915.
- [13] Sarker, Abeed, et al. "Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from Twitter." Drug safety 39.3 (2016): 231-240.
- [14] Chan, Brian, Andrea Lopez, and Urmimala Sarkar. "The canary in the coal mine tweets: Social media reveals public perceptions of non-medical use of opioids." PloS one 10.8 (2015): e0135072.
- [15] Graves, Rachel L., et al. "Opioid Discussion in the Twittersphere." Substance use & misuse (2018): 1-8.
- [16] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;12:2825–2830.
- [17] Chollet F. Keras. GitHub; 2015.
- [18] Pre-trained word2vec model; <https://radimrehurek.com/gensim/models/word2vec.html>