Detecting Illicit Drug Ads in Google+ Using Machine Learning

Fengpan Zhao
Georgia State University, fengpan.zhao@outlook.com

Pavel Skums
Georgia State University, pskums@gsu.edu

Alex Zelikovsky
Georgia State University, azelikovsky@cs.gsu.edu

Eric L. Sevigny
Georgia State University, esevigny@gsu.edu

Monica Swahn
Georgia State University, mswahn@gsu.edu

See next page for additional authors

Authors
Fengpan Zhao, Pavel Skums, Alex Zelikovsky, Eric L. Sevigny, Monica Swahn, Sheryl M. Strasser, and Yubao Wu

Detecting Illicit Drug Ads in Google+ Using Machine Learning

Fengpan Zhao[1,5][0000−0003−1355−1330], Pavel Skums[1,5], Alex Zelikovsky[1,5], Eric L. Sevigny[3,5], Monica Haavisto Swahn[4,5], Sheryl M Strasser[4,5], and Yubao Wu[1,5]

[1] Department of Computer Science, Georgia State University, Atlanta, GA, USA
[2] Department of Criminal Justice and Criminology, Georgia State University, Atlanta, GA, USA
[3] School of Public Health, Georgia State University, Atlanta, GA, USA
[4] email: fzhao6@student.gsu.edu, pskums@gsu.edu, alexz@cs.gsu.edu, esevigny@gsu.edu, mswahn@gsu.edu, sstrasser@gsu.edu, ywu28@gsu.edu

Abstract. Opioid abuse epidemics is a major public health emergency in the US. Social media platforms have facilitated illicit drug trading, with significant amount of drug advertisement and selling being carried out online. In order to understand dynamics of drug abuse epidemics and design efficient public health interventions, it is essential to extract and analyze data from online drug markets. In this paper, we present a computational framework for automatic detection of illicit drug ads in social media, with Google+ being used for a proof-of-concept. The proposed SVM- and CNN-based methods have been extensively validated on the large dataset containing millions of posts collected using Google+ API. Experimental results demonstrate that our methods can efficiently identify illicit drug ads with high accuracy. Both approaches have been extensively validated using the dataset containing millions of posts collected using Google+ API. Experimental results demonstrate that both methods allow for accurate identification of illicit drug ads.

Keywords: Illicit drug ads · social media · text mining · deep learning

1 Introduction

The opioid abuse epidemic is a national crisis seriously affecting public health, causing preventable harm and premature death, and devastating communities. In 2017, 70,467 Americans died of drug overdoses that year, representing an increase of 10 percent over the 63,938 opioid overdose deaths recorded in 2016 [1].

The drug abuse epidemic has been facilitated by modern information technology and the rise of illicit drug trading platforms. With an estimated 4.1 billion persons worldwide regularly using the Internet in 2018 [2], drug vendors can efficiently and effectively reach drug consumers via online social media platforms. Online drug trading is both more efficient and less risky than traditional drug market exchanges, since the buyer does not need to connect with the seller in person. An open question concerns the ectent to which the current opioid abuse epidemic is facilitated by the proliferation of social media. In our research, we found that most social media platforms are used extensivelyfor illicit drug advertising. Figure 1 shows two sample posts collected from Google+. Many ads contain vendors' phone numbers, emails, Wickr IDs, and websites. Buyers can contact drug vendors using these communication methods to order drugs for delivery to a specified pickup location. Purchasing illicit drugs online seemingly has become as straightforward as making an Amazon purchase. It is therefore of paramount importance that public health and law enforcement personnel have access to efficient tools for monitoring online drug transactions using traditional epidemiological surveillance methods to inform the design of appropriate response strategies.

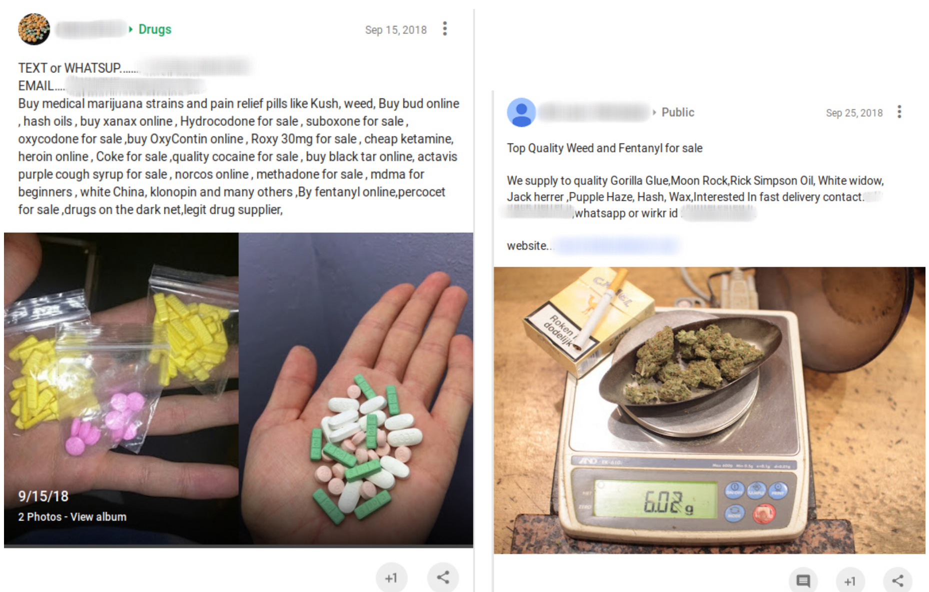

Fig. 1. Examples of illicit drug advertisements from Google+

In this paper, we develop a computational framework for detecting illicit ads in Google+, one of the largest social media platforms. We first captured relevant data posts via Google+ APIs, and then applied binary classification methods to analyze the text data in the posts. The textual analyses were used to identify illicit drug ads. We employed two methods in our approach: 1. the support vector machine (SVM) and 2. the convolutional neural network (CNN). The SVM-based method allowed for term frequency-inverse document frequency (TF-IDF) extraction of terms, which were subsequently applied to SVM for prediction [3, 4]. The CNN-based methodology was applied to social media posts for text classification [5]. The first approach (SVM) required a precursory feature selection, while the second approach (CNN) automatically learns features from the text data.

2 Related Work

Illicit online drug trade has been the subject of several epidemiological and sociological studies. In particular, Mackey et al. [6] created a fictitious advertisement, offering consumers a way to buy drugs without a prescription. The advertisement was posted on four social media platforms: Facebook, Twitter, MySpace and Google+. Eventually one of these accounts was blocked due to suspicious activity, but the remaining fake illicit drug advertisements were easily accessible during the duration of the experiment. A study conducted by Stroppa et al. [7] revealed that one-fifth of collected posts advertise counterfeit and/or illicit products online. Their research emphasized that detection of illegal cyber-vendors and online tactics requires development and application of sophisticated and tailored screening/detection methods.

On the computational side, development of tools for detection of malicious and/or undesired advertisements in social media has been a subject of several studies. Hu et al. [8] provided a framework for detection of spammers on microblogging. Zheng and colleagues [9] proposed a SVM-based machine learning model to detect spammer behavior on Sina Weibo. Agrawal et. al [10] introduced an unsupervised method called Reliability-based Stochastic Approach for Link-Structure Analysis, which can be used to detect topical posts on social media. Jain et al. [11] used convolutional and long short-term memory (LSTM) neural networks to detect spam in social media, while addressing the challenges of text mining on short posts.

In contrast to the previous studies, we specifically focus on detection of illicit drug ads within social media platforms, with the aim of applying epidemiological methods to investigate online enabling structures associated with opioid abuse.

3 Methods

In this section, we describe two methods for classifying social media posts based on Support Vector Machine (SVM) and Convolutional Neural Network (CNN) approaches. For both methods, the inputs are text data extracted from Google+ posts, and the outputs are the predicted labels indicating whether each post is an illicit drug ad.

3.1 The SVM-based Method

The proposed method pipeline consists of two stages: pre-processing and classification.

Pre-processing Steps. At this stage, text posts collected from social media are transformed into numerical feature vectors, which are further used as the inputs for the SVM classifier. It is a crucial part of traditional text mining methods because the selected features affect the performance of the classifier. Figure 2 shows the general scheme of the pre-processing stage.

Fig. 2. Pre-processing steps

Pre-processing consists of three steps. In the first step, the stop words considered noise are removed. In the second step, the root of a word is isolated by removing tenses of verbs, which is also referred to as stemming [12]. In the third step, the term frequency-inverse document frequency (TF-IDF) features are determined [13]. The TF-IDF is the product of two statistics: term-frequency and inverse document frequency. The term frequency is calculated based on the raw count of a term (word). The inverse document frequency is a measure of how much information the word provides.

Support Vector Machine (SVM) Classification . TF-IDF features computed at the pre-processing step are used to train an SVM model that can be further used to predict labels of new posts. SVM is a classical supervised learning method, which constructs a hyperplane in a multidimensional euclidean space to serve as a separator for feature vectors from two classes. We used the radial basis function (RBF) kernel SVM classifier, whose accuracy was assessed using a ten-fold cross-validation process on a labeled post text dataset manually curated by human experts.

3.2 The CNN-Based Method

This method uses the TextCNN approach [5], which first computes a word embedding and then applies the convolutional neural networks (CNN) to perform the classification. TextCNN does not require the removal of stop words or stemming.

Word embedding. Word embedding which maps words or phrases to numerical vectors, was utilized to allow neural networks to process the text data. We used Word2vec, a commonly used word embedding model [14] that relies on the combination of skip-grams and continuous bag-of-words (CBOW) procedures [15]. CBOW generates a word based on the context, while skip-grams generates the context from a word. For example, if we treat {"Washington D.C.", "is", "the United States"} as a context, then CBOW will generate the word "capital". If given the word "capital", skip-grams will be able to predict the following words: 'Washington D.C.", "is", "the United States". The numerical vectors generated by word2vec are used as the input of CNN.

Convolutional Neural Networks. TextCNN contains a single layer of neural net, which allows it to be highly scalable yet sensitive in performing text classification. Figure 3 shows the general scheme of TextCNN[16]. Let d be the dimension of word vector. Given a sentence "Buy drugs on social media without prescription" and $d = 5$, we can generate a sentence matrix in Figure 3. Then feature maps are generated by filters operating convolutions on the sentence matrix. Here we set the region sizes to 2, 3 and 4, and each region size has two filters. A max-pooling operation is applied to the feature map to retrieve the largest number. Therefore we can take six features from six feature maps and concatenate them together to get a feature vector which will serve as the input of the softmax layer. Finally, we complete a binary classification by using this feature vector through softmax layer.

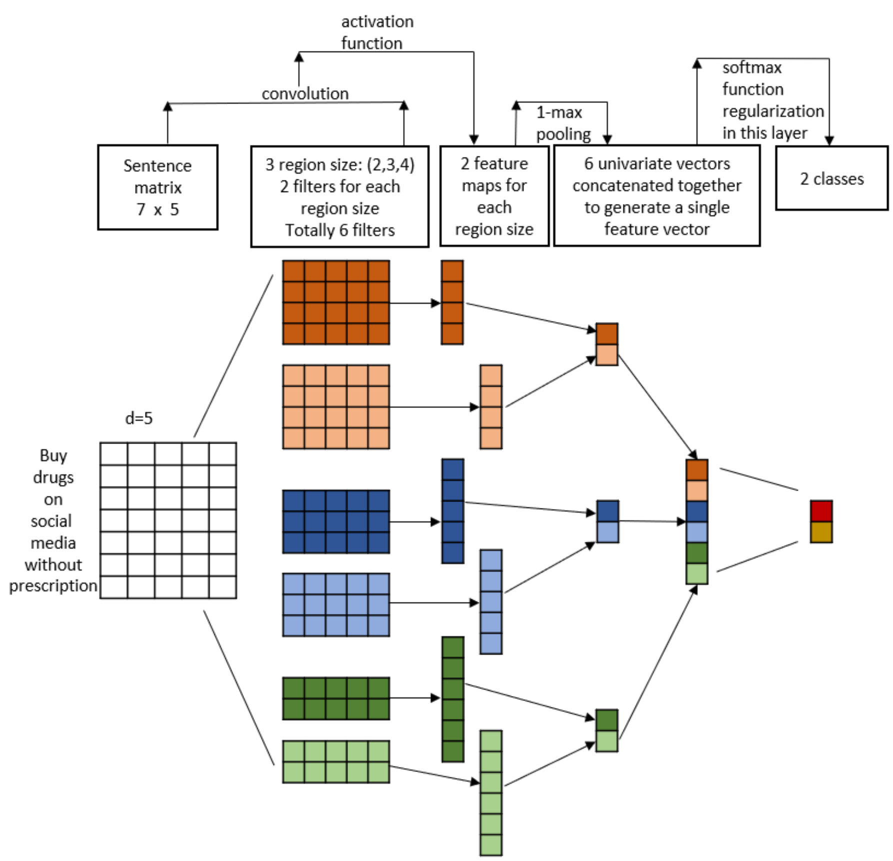

Fig. 3. Illustration of TextCNN

4 Experimental Results

In this section, we will describe the data collection and data processing, and then evaluate the performance of the SVM-based and CNN-based methods. All tools have been implemented in Python 2.7, and run on a DELL workstation with Intel Xeon E5-1603 2.80GHz CPU, 32G memory, and Ubuntu 18.04 OS.

4.1 Data Collection

The data have been collected using Google+ API. The analyzed dataset has been formed by posts containing at least one of the following 30 keywords [17]:

opioid, alprazolam, amphetamine, antidepressant, benzodiazepine, buprenorphine, cocaine, diazepam, fentanyl, heroin, hydrocodone, meth, methadone, morphine, naloxone, narcan, opana, opiate, overdose, oxycodone, oxymorphone, percocet, suboxone, subutex, pill, rehab, sober, withdrawal, shooting up, track marks

In total, 1,162,445 posts published from 2018/01/01 to 2018/10/31 have been collected. We labeled all the posts manually. The following examples illustrate examples of illicit drug ads from the dataset. Ads 1-3 are selling illicit drugs while ad 4 is a normal post.

1. Buy pain pills and other research chemicals. We do offer discount as well to bulk buyers. Overnight Shipping with tracking numbers provided. Stay to enjoy our services.Overnight shipping with a tracking number provided for your shipment(Fast,safe and reliable delivery). We ship within USA, AUSTRALIA, CANADA, GERMANY, POLAND, SWEDEN, NEW ZEALAND and many other countries not listed here.
2. Hello we supply high quality medication and high rated pharmaceutical opioid at affordable prices. Dear buyers we bring you The Best Of real pharmaceutical product such as oxycodone, nembutal powder, fentanyl patch and fentanyl powder, subutex, adderal, demerol, hydrocodone MDMA etc, and only serious buyers should contact please.
3. Hello, I am a vendor in high quality pharmaceutical products like Xanax, Oxycodone, Fentanyl patch, Viagra, Diazapam, Percoset, Opana, Methadone, etc and also high quality medical marijuana strains like Og kush, Sativa, Kief,S hatter, Girls Scott, Lemon haze, Moon rock, Afghan kush, Purple haze etc, my packaging is very safe and discreet, also my delivery is 100% assured as we do refund or resend the same order immediately in case of any unforeseen.
4. Highlighting concerns with the pharmaceutical supply chain, the Food and Drug Administration warned McKesson, one of the nations largest wholesalers, for failing to properly handle episodes where pharmacies received tampered medicines, including three ...
   FDA scolds McKesson for naproxen in tampered oxycodone bottles -STAT-

4.2 Effectiveness Evaluation

We use precision, recall and F-score as metrics to evaluate the accuracy of the classification methods [18]. Precision is defined as the ratio of predicted and ground-truth illicit ads among all predicted illicit ads, i.e., $\text{Prec} = \text{tp}/(\text{tp} + \text{fp})$. Recall is defined as the ratio of predicted and ground-truth illicit ads among all ground-truth illicit ads, i.e., $\text{Recall} = \text{tp}/(\text{tp} + \text{fn})$. The F-score is the harmonic mean of precision and recall: $\text{F-score} = 2 \cdot \text{Prec} \cdot \text{Recall}/(\text{Prec} + \text{Recall})$. We use 10-fold cross-validation procedures to evaluate the accuracy of both the SVM and CNN based methods.

In TextCNN, we set the parameters as follows: max_sequence_length 20, embedding_dim 200, validation_split 0.16, test_split 0.2 [16]. Table 1 shows the precision, recall, and F-score for SVM and TextCNN. From Table 1, we can see that TextCNN outperforms SVM in all metrics.

Table 1. Accuracy of the SVM based method and TextCNN

Methods	Pre	Recall	F-score
SVM Based Method	0.65	0.81	0.72
TextCNN	0.97	0.90	0.93

Table 2 shows the running time. In Table 2, the training time represents the average running times for training ten SVM or CNN models during the ten-fold cross-validation procedure. The number of posts in the input dataset for training each model is 1,046,200, which is 90% of the total of 1,162,445 posts. The testing time represents the average running time of predicting the label of a single post. In each iteration of the ten-fold cross-validation, the input number of posts is 116,244 posts. We measure the average time for each post. From Table 2, we can see that the SVM based method takes less than 1 hour while the TextCNN method takes 11 hours for training. Both methods take less than 0.05 second for prediction.

Table 2. Running time of the SVM based method and TextCNN

Methods	Training time	Testing time
SVM Based Method	2,469s	0.023s
TextCNN	3,936 s/epoch, 10 epoch	0.034s

5 Conclusion

Social media platforms have facilitated illicit drug trading and may be an important driver of the current opioid epidemic. Thus tools for monitoring and analysis of online drug markets are needed to advance epidemiological studies and develop intervention applications. In this paper, we used the Google+ platform as a proof-of-concept to demonstrate that machine-learning-based methods allow for efficient identification of illicit drug advertisements from social media posts. Our tools could be used by health care practitioners, law enforcement officials and researchers to extract and analyze the data relted to the opioid abuse epidemic, which can be examined to better understand dynamics of online drug markets, trade, and behaviors. These insights are essential in the development of tailored recommendations and public health intervention strategies that are responsive to social media and online environments.

References

1. Centers for Disease Control and Prevention: Provisional Drug Overdose Death Counts. https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm. Accessed Febrary 3, 2019
2. Stevens, J.: Internet Stats & Facts for 2019. https://hostingfacts.com/internet-facts-stats/. Accessed December 17, 2018
3. Wu, H.C., Luk, R.W.P., Wong, K.F., Kwok, K.L.: Interpreting tf-idf term weights as making relevance decisions. ACM Transactions on Information Systems (TOIS) **26**(3), 13 (2008)
4. Suykens, J.A., Vandewalle, J.: Least squares support vector machine classifiers. Neural processing letters **9**(3), 293–300 (1999)
5. Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882 (2014)
6. Mackey, T.K., Liang, B.A.: Global reach of direct-to-consumer advertising using social media for illicit online drug sales. Journal of medical Internet research **15**(5) (2013)
7. Stroppa, A., di Stefano, D., Parrella, B.: Social media and luxury goods counterfeit: a growing concern for government, industry and consumers worldwide. The Washington Post (2016)
8. Hu, X., Tang, J., Zhang, Y., Liu, H.: Social spammer detection in microblogging. In: Twenty-Third International Joint Conference on Artificial Intelligence (2013)
9. Zheng, X., Zeng, Z., Chen, Z., Yu, Y., Rong, C.: Detecting spammers on social networks. Neurocomputing **159**, 27–34 (2015)
10. Agrawal, M., Velusamy, R.L.: R-salsa: A spam filtering technique for social networking sites. In: 2016 IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1–7 (2016). IEEE
11. Jain, G., Sharma, M., Agarwal, B.: Spam detection in social media using convolutional and long short term memory neural network. Annals of Mathematics and Artificial Intelligence **85**(1), 21–44 (2019)
12. Hull, D.A.: Stemming algorithms: A case study for detailed evaluation. Journal of the American Society for Information Science **47**(1), 70–84 (1996)
13. Salton, G., McGill, M.J.: Introduction to modern information retrieval (1986)
14. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
15. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, pp. 3111–3119 (2013)
16. Zhang, Y., Wallace, B.: A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. arXiv preprint arXiv:1510.03820 (2015)
17. Wu, Y., Skums, P., Zelikovsky, A., Rendon, D.C., Liao, X.: Predicting opioid epidemic by using twitter data. In: International Symposium on Bioinformatics Research and Applications, pp. 314–318 (2018). Springer

18. Powers, D.M.: Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation (2011)