Drug Abuse Detection via Broad Learning

Chao Kong[1(✉)], Jianye Liu[1], Hao Li[1], Ying Liu[1], Haibei Zhu[2], and Tao Liu[1]

[1] School of Computer and Information, Anhui Polytechnic University, Wuhu, China
kongchao315@163.com, 17855368016@163.com, lhthomas@163.com,
liuyinglyj_011@163.com, liutao@ahpu.edu.cn
[2] School of Electrical Engineering, Anhui Polytechnic University, Wuhu, China
zhb877097717@163.com

Abstract. Prescription drug abuse is one of the fastest growing public health problems in the USA. This work develops a broad learning method for Drug Abuse Detection (DAD). In this paper, we propose a new broad learning-based method named ILSTM, short for *Improved Long Short-Term Memory*, to study the data fusion and prediction from heterogeneous data sources for DAD. The algorithm utilizes the broad learning framework to handle data fusion broadly and information mining deeply simultaneously. Moreover, the effectiveness and prevalence of Holt-Winter inspire our work in the temporal property for DAD.

Keywords: Broad learning · Data fusion · Drug Abuse Detection

1 Introduction

The United States is experiencing a national crisis regarding the use of synthetic and nonsynthetic opioids, either for the treatment and management of pain (legal, prescription use) or for a recreational purpose (illegal, non-prescription use). Federal organizations such as the Centers for Disease Control (CDC) are struggling to save lives and prevent negative health effects of this epidemic, such as opioid use disorder, hepatitis, HIV infections and neonatal abstinence syndrome."[1] To address this epidemic, we develop a broad learning method for Drug Abuse Detection (DAD). The broad learning method is an ubiquitous model to achieve better learning or mining performance on solving the real-world problem in the era of big data, which all kinds of data are available. It has been widely used in many applications such as POI recommendation [1], link prediction [2], fraud detection [3] and so on.

Generally speaking, DAD is the problem of monitoring the illegal drug or prescription medication abuse, predicting drug abuse trends, classifying people who may be caught in drug abuse or not. To perform predictive analytics on DAD, it is crucial to first fuse or integrate multiple available heterogeneous data

---

[1] http://www.cdc.gov/features/confronting-opioids/index.html.

sources. However, existing works have primarily focused on the single data source such as Tweets [4]. Following the pioneering work of [4], these methods typically apply a three-step solution: first collecting drug abuse-related tweets at a large-scale, then designing an annotation strategy (drug abuse vs. non-drug abuse), and last developing a deep learning model that can accurately classify tweets into drug abuse risk behavior.

To the best of our knowledge, none of the existing works has paid special attention to DAD with broad learning framework. In this work, we focus on the problem of broad learning for DAD. We propose ILTSM (short for *Improved Long Short-Term Memory*), which addresses the aforementioned limitations of existing works. Below we highlight our major contributions in this work.

– To account for both the explicit information and implicit associations (spatio-temporal information and socio-economic data), we propose a new broad learning framework. In particular, the algorithm can handle data fusion broadly and information mining deeply simultaneously.
– To address the characteristic of time-lag, we set the dual gate to update new and effective data, forget old or invalid data in the pre-processing phase. Moreover, we employ Holt-Winter to predict the drug abuse trends in the time dimension, which can revise the prediction curve to denoise.

The rest of paper is organized as follows. We shortly discuss the related work in Sect. 2. We present the DAD method in Sect. 3 and report our empirical study in Sect. 4. Finally, we conclude this paper in Sect. 5.

2 Related Work

Drug abuse detection aims at predicting prevalence and patterns of abuse of both illegal drugs and prescription medications. The study of DAD problem has become a hot topic in recent years, and some earlier studies can go back to 2009s [5]. As an ubiquitous social media, Twitter is one of the most popular social networks, with more than 115 million monthly active users and over 58 million tweets per day.[2] Twitter is currently being used as a major resource in various detection tasks, including discrimination detection [6], influenza epidemic discovery [5], sentiment analysis [7], drug abuse detection, sexual health monitoring, and pharmacovigilance. To date, the existing works have primarily focused on the detection and analysis of illicit and prescription drug abuse using tweets. In general, the existing studies of DAD can be mainly divided into two categories: automatic monitoring and bag-of-words model. The former employs machine learning methods for an automatic classification that can identify tweets which are indicative of drug abuse [8]. Chary et al. [8] discussed how to use artificial intelligence techniques to extract content useful for purposes of toxicovigilance from social networks. However, the latter employs the bag-of-words model to build a dictionary, computes the similarity of the data items with a probabilistic manner and predicts the drug abuse trends based on a proposed decision or

---

[2] http://www.statisticbrain.com/twitter-statistics/.

score function [9]. Traditional DAD methods are developed on single social data without considering the data fusion of spatio-temporal information and socio-economic data. So it is difficult to guarantee better learning and optimization performance. These are the focuses of this work.

3   Drug Abuse Detection Approach

3.1   Overview of ILSTM

Our proposed approach consists of three components as following.

Step1. Feature selection via CNN. CNN can learn and provide the mapping between input and output, which does not require any precise mathematical expression between the input and output. The input features are converted to a two-dimensional matrix. The training phase compresses it to obtain the actual features, enabling the convolutional neural networks to map the input data into eigenvalue accurately.

Step2. Data fusion via ILSTM. Due to the characteristic of time-lag in datasets, we try to employ LSTM's [10] improved variant ILSTM to train. First, we set a dual gate to determine what information should be discarded from memory state: $f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f)$, where W_f and b_f represent weight and bias of **sigmoid** activation function respectively. Then, the sigmoid layer of the input gate layer determines which information needs to be updated by **tanh** layer. Finally, the ILSTM unit controls the output information: $o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o)$, $h_t = o_t * tanh(C_t)$. For ILSTM model, we treat the output layer of CNN as the input layer of bidirectional GRU to perform the feature extraction and data fusion.

Step3. Multi-class classification. We perform the information fusion and prediction by employing bidirectional ILSTM, and the predicted results are learned forward in full connection layer. The connection parameter batch is denoted as o^t and h^t respectively in the final fusion layer. Then, we make the derivation of weight W and biases coefficient b. Finally, we normalize the output layer by softmax function and give the probability distribution of all the attribute values: $P_r(a) = \frac{\exp(o_{t_i})}{\sum_{i=1}^{n} \exp(o_{t_i})}$.

3.2   Parameters Inference and Prediction

For each dataset, we define the attribute feature matrix as A_f. We put N attribute feature matrices $A_{f_i}(1 < i \leq N)$ into the input layer and extract the information through B-GRU [11]: $\overrightarrow{H}(A_{f_i}) = \overrightarrow{GRU}(A_{f_i})$, $\overleftarrow{H}(A_{f_i}) = \overleftarrow{GRU}(A_{f_i})$, $H(A_{f_i}) = \overrightarrow{H}(A_{f_i}) \times \overleftarrow{H}(A_{f_i})$, where $\overleftarrow{H}(A_{f_i})$ and $\overrightarrow{H}(A_{f_i})$ represent the information extracted by performing forward GRU and backward GRU respectively. When the last fusion layer extracted the information by ILSTM, $\overrightarrow{h_{i_l}}^t$ and $\overleftarrow{h_{i_l}}^t$

represent the information from ILSTM to full connection layer to perform a multi-class classification task.

First, we take the partial derivatives of output o^i : $\frac{\theta o^i}{\theta h_{ij}} = \frac{\sum_{j=1}^{length(input)} w_{ij} * h_{ij}}{\theta x_j} = w_{ij}$, where a_i represents the output of the i-th layer, and w_{ij} denotes the weight of j-th input in i-th layer. Then, we take the partial derivatives of loss:

$$\frac{\theta loss}{\theta h_{ij}} = \sum_{j}^{length(output)} \frac{\theta loss}{\theta o^i} \frac{\theta o^i}{\theta h_{ij}} = \sum_{j}^{length(output)} \frac{\theta loss}{\theta o^i} * w_{ij}, \tag{1}$$

which proves the backpropagation from $(i+1)$-th layer to i-th layer. Next, we can derive the weight w by

$$\frac{\theta loss}{\theta w_{h_{il}}} = \frac{\theta loss}{\theta o^i} \frac{\theta o^i}{\theta w_{h_{il}}} = \frac{\theta loss}{\theta o^i} * h_{ij} \tag{2}$$

The output O^t from full-connection layer is: $O^t = w_{\overrightarrow{h_{i_l}}}\overrightarrow{h_{i_l}} + w_{\overleftarrow{h_{i_l}}}\overleftarrow{h_{i_l}}$, where $w_{\overrightarrow{h_{i_l}}}$ and $w_{\overleftarrow{h_{i_l}}}$ denote the weight of forward pass and backward pass respectively. Once parameters O^t is estimated, ILSTM can classify each class V_{R_i} by computing its probability. Due to the page limitation, we omit the proofs and computations. To select the computation of probabilities significantly, here we employ the softmax function:

$$V_{R_i} = \frac{e^{\sum_{t=0}^{T_k-1} O^{t_k}}}{\sum_{i=0}^{C-1} e^{\sum_{t=0}^{T_i-1} O^{t_i}}}, \tag{3}$$

where C means the classes of each feature. Finally, we choose the *top*-k highest probability to predict the counties which most likely outbreak the drug abuse at a certain time.

4   Empirical Evaluation

We crawled three datasets from MCM/ICM contest 2019[3] and Twitter[4] for our experiments. The *ACS* covers socio-economic factor in each county. The *DEA* describes the number of opioid abuse reports of counties in each year. *Twitter* collects the comments from Twitter' users about drug abuse. The descriptive statistics about the datasets are shown in Table 1. We find a prevalent DAD's solution, called Support Vector Machine (shorted in SVM), to be the comparative baseline. As mentioned above, we evaluate our proposed method using *Precision* and *Recall*.

We extracted drug abuse reports, geographic location information, age, education and other socio-economic information from Twitter to integrate DEA

---

[3] https://www.comap.com/undergraduate/contests/.
[4] http://followthehashtag.com/datasets/free-twitter-dataset-usa-200000-free-usa-tweets/.

Table 1. Descriptive statistics of datasets

Data source	#records	#features
DEA	24063	10
ACS	3245	151
Twitter	204821	15

(a) Precison

(b) Recall

(c) Precision with 50% noise

(d) Recall with 50% noise

(e) ILSTM

(f) SVM
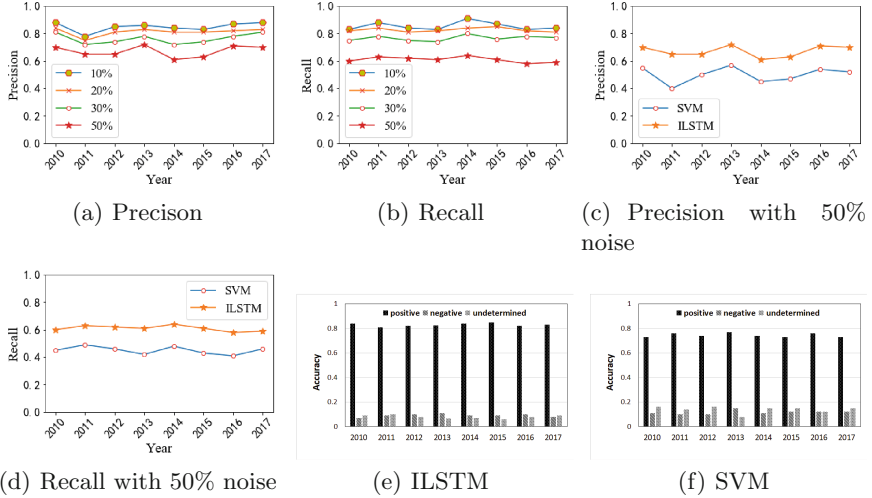
Fig. 1. Performance comparison of DAD on DEA, ACS and Twitter

and ACS to make the prediction. Moreover, Holt-Winter [12] was employed to smooth the output and remove noise. Figures 1(a) and (b) manifest the robustness of ILSTM by injecting noise whose ratio is varied from 10% to 50%. We observe that the accuracy of ILSTM is promised. If 10% noise is injected into data, almost 90% of drug abuse counties can be found by ILSTM in 2010. Even 50% noise is injected into the data, the precision in 2010 is almost 70%. Figures 1(c) and (d) illustrate that ILSTM outperforms the baseline significantly. We now turn to DAD on three heterogeneous data sources without noise namely clean DEA, ACS and Twitter. We manually annotated the accuracy from 2010 to 2017 labeled by ILSTM and SVM. As shown in Figs. 1(e) and (f), we find that the accuracy for the DAD of ILSTM from 2010 to 2017 is more than 80%, but it is about 65% for SVM. This illustrates that both ILSTM and SVM are quite good in returning the correct drug abuse counties from 2010 to 2017. ILSTM also returns fewer undetermined results than SVM. In summary, the result indicates that implicit associations are helpful to achieve better performance for DAD.

5   Conclusion

In this paper, we have studied the problem of Drug Abuse Detection via broad learning across two kinds of heterogeneous data sources. It is a challenging task due to the characteristic of time-lag, implicit associations and data fusion. We propose a supervised method to deal with the mentioned challenges. We have illustrated our proposed method on three real data sources. Experimental results demonstrate the effectiveness and rationality of our ILSTM method. In the future, we plan to expand the datasets to insert more explicit and implicit features, make the predictive effect of DAD closer to the truth and assess the impact of more drugs on DAD in target areas.

References

1. Wang, F., Qu, Y., Zheng, L., Lu, C.-T., Yu, P.S.: Deep and broad learning on content-aware POI recommendation. In: CIC 2017, USA, pp. 369–378 (2017)
2. Zhang, J., Xia, C., Zhang, C., Cui, L., Fu, Y., Yu, P.S.: BL-MNE: emerging heterogeneous social network embedding through broad learning with aligned autoencoder. In: ICDM 2017, USA, pp. 605–614 (2017)
3. Cao, B., Mao, M., Viidu, S., Yu, P.S.: HitFraud: a broad learning approach for collective fraud detection in heterogeneous information networks. In: ICDM 2017, USA, pp. 769–774 (2017)
4. Hu, H., et al.: Deep learning model for classifying drug abuse risk behavior in tweets. In: ICHI 2018, USA, pp. 386–387 (2018)
5. Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L.: Detecting influenza epidemics using search engine query data. Nature **457**(7232), 1012–1014 (2009)
6. Wu, Y., Wu, X.: Using loglinear model for discrimination discovery and prevention. In: DSAA 2016, Canada, pp. 110–119 (2016)
7. Keith Norambuena, B., Lettura, E.F., Villegas, C.M.: Sentiment analysis and opinion mining applied to scientific paper reviews. Intell. Data Anal. **23**(1), 191–214 (2019)
8. Chary, M., Genes, N., McKenzie, A., Manini, A.F.: Leveraging social networks for toxicovigilance. J. Med. Toxicol. **9**(2), 184–191 (2013)
9. Balsamo, D., Bajardi, P., et al.: Firsthand opiates abuse on social media: monitoring geospatial patterns of interest through a digital cohort. In: WWW 2019, USA, pp. 1–7 (2019)
10. Han, X., Xu, L., Qiao, F.: CNN-BiLSTM-CRF model for term extraction in Chinese corpus. In: Meng, X., Li, R., Wang, K., Niu, B., Wang, X., Zhao, G. (eds.) WISA 2018. LNCS, vol. 11242, pp. 267–274. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-02934-0_25

11. Wang, S., Wu, B., Wang, B., Tong, X.: Complaint classification using hybrid-attention GRU neural network. In: Yang, Q., Zhou, Z.-H., Gong, Z., Zhang, M.-L., Huang, S.-J. (eds.) PAKDD 2019. LNCS (LNAI), vol. 11439, pp. 251–262. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-16148-4_20

12. Raikwar, A.R., et al.: Long-Term and short-term traffic forecasting using holt-winters method: a comparability approach with comparable data in multiple seasons. IJSE **8**(2), 38–50 (2017)