Distilling Meta Knowledge on Heterogeneous Graph for Illicit Drug Trafficker Detection on Social Media - Supplementary Material

Anonymous Author(s)
Affiliation
Address
email

In the supplementary material, to help better understand our model, we introduce the details of data preparation, feature and relation extraction, as well as the pseudo-code of MetaHG. To help reproduce our model and all baseline models, we provide a detailed description of experimental settings (including experimental environment, the value of hyper-parameters, iteration number, and details of reproducing baseline models). Additionally, we conduct additional experiments to validate the robustness of MetaHG and discuss the potential ethical issues and the limitation of our paper.

1 Dataset Details

1.1 Data Preparation

Based on the official Instagram APIs [1], we collect 8,651 users and 79,705 posts from Instagram. According to the drug types defined by National Institute on Drug Abuse [8], we manually classify these users into six groups (including five types of drug traffickers and regular users) based on the functions of drugs they post on Instagram (see Table 1). In Table 1, the left column represents the drug trafficker group and the right column shows the corresponding drugs belonging to the group. Regular users are those who are irrelevant to drug trafficking activities on Instagram. For instance, hallucinogens traffickers are defined as those who only sell hallucinogens-related drugs (e.g., LSD and DMT). Note that mixture traffickers are those who sell at least two groups of drugs on Instagram. For instance, a drug trafficker who sells depressant-related drugs (e.g., xanax and alprzaolam) and opioid-related drugs (e.g., fentanyl and oxycodone) is defined as a mixture drug trafficker. In this paper, according to the number of labeled samples, we consider three types of drug traffickers (stimulants traffickers, depressants traffickers, mixture traffickers) as training task data and the rest of two types (opioids traffickers and hallucinogens traffickers) as testing task data.

Table 1: The different types of drug traffickers and their related drugs.

Trafficker Type	Drugs
Stimulants trafficker	cocaine, meth (crystal meth), amphetamine, methamphetamine, weed
Depressants trafficker	xanax, fermapram, valium, halcion, ativan, klonopin, alprzaolam
Hallucinogens trafficker	LSD, MDT, MDMA, ketamine, magic mushrooms, mescaline, hoasca
Opioids trafficker	oxycodone, hydrocodone, codeine, morphine, fentanyl, meperidine
Mixture trafficker	sell at least two different groups of drugs (e.g., cocaine, xanax, and LSD)

1.2 Feature and Relation

In this paper, we employ multi-modal features (including text and image features of users and posts) and structure information (eight types of relations) among nodes to build the HG.

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

Content Feature. Most users on social media post images and text simultaneously. Thus, we consider both text feature and image feature of posts and users. ***Text Feature:*** we first merge all of text information as corpus to pre-train language model *BERT* [2] and convert all of the text information for each node to a fixed-length feature vector ($d = 200$). Specifically, for keyword feature, we extract all of the keywords from every post and user profile to represent the node. Then, we select a set of illicit-oriented keywords based on word frequency and feed the keyword sets to *BERT* to obtain the embedding of keyword node. For enumerated attributes (e.g., # of posts/followings/followers), we apply one-hot encoding to convert it to a binary feature vector. Finally, for each node, we concatenate all pre-trained features as the text feature vector. ***Image Feature:*** we employ the pre-trained image model *VGG19* [12] to acquire the embedding ($d = 1000$) of each image and then implement *PCA* [10] to decrease the dimension from 1000 to 200. For keyword nodes, the image feature for each keyword node is set as zero. Finally, both text and image features are concatenated as the attribute feature vector for each node.

Relation. To determine whether a user is a drug trafficker on social media, we not only consider the content-based features (text and image features), but also the complex relationships among users, posts, and keywords. To characterize the relatedness of two nodes, we consider eight kinds of relationships as follows. *R1: user-follow/followed-user* denotes that a user is following or followed by another user. *R2: user-tagger-user* denotes that a user tags another user in some posts. *R3: user-reply-post* denotes that a user replies to a post of another user. *R4: user-mention-post* denotes that a user mentions another user in a post. *R5: user-have-post* denotes that a post belongs to the user. *R6: user-profile-keyword* denotes that the profile description of a user contains the keyword. *R7: post-include-keyword* denotes that the post content includes the keyword. *R8: post-tag-keyword* denotes that the post content has the hashtag keyword.

2 Algorithm Details

The pseudo-code of MetaHG training procedure is shown as follows:

Algorithm 1 Training Procedure of MetaHG

Require X, A, ϕ: nodes features, adjacent matrix, and randomly initialized parameters
1: Learn a refined graph via Equation 3.
2: Implement self-supervised learning to augment R-GCN on the refined graph.
3: **while** not convergae **do**
4:     Sample a batch of training tasks τ from $\mathcal{T}$.
5:     **for** each τ **do**
6:         Sample a support set $\mathcal{S}_\tau$ and a query set $\mathcal{Q}_\tau$.
7:         Update parameters $\phi_\tau^{'}$ via Equation 8.
8:     **end for**
9:     Sample a support set $\mathcal{S}_\tau$ and a query set $\mathcal{Q}_\tau$ from meta-testing task.
10:     Distill the soft knowledge from the teacher model using $\mathcal{Q}_\tau$ via Equation 10.
11:     Train the student model via optimizing Equation 12 and update meta-testing model parameters ϕ^* via Equation 13.
12: **end while**
13: **Return** Optimized ϕ^*

3 Experimental Details

3.1 Baseline Setting

We employ five sets of baseline models (twenty) in this paper. To compare with few-shot learning models fairly, in traditional classifiers (B1, B2, and B4), we utilize all labeled data of the training tasks and few-shot labeled data (support set) of the testing tasks for model training and then we use the rest of data of testing tasks (query set) for model evaluation. For B1 group, we take text feature (tFeature), image feature (iFeature), and the combination of text and image features (cFeature) as features vectors for users respectively and feed them into a 3-layer **DNN** [14] classifier to detect drug traffickers. For B2 group, we reproduce the method [6] by implementing a recurrent neural network with an LSTM unit to study the pattern of users with tFeature to detect drug traffickers. Additionally, we reproduce the model [7] by implementing the biterm topic model with tFeature to learn the latent patterns of users for detecting drug traffickers. For B4 group, we implement **DeepWalk** [9] to learn node embedding (ignoring the heterogeneous property and attribute information) by modeling structure proximity. Besides, we implement **metapath2vec** [3] to learn the semantic information of defined meta-paths [15] in this application. In addition, we also implement four graph neural network based representation learning models including **GCN** [5], **GAT** [16], **HAN** [18], and **R-GCN** [11] to learn the node embedding in HG by leveraging both node feature and graph structure information. In particular, for HAN, we utilize the defined meta-paths and implement HAN to learn the attention-based node embedding. Similar to B1 and B2, we feed the learned user embedding to a 3-layer DNN classifier to detect illicit drug traffickers.

For few-shot learning classifiers (B3, B5), we define few-shot samples in each task as the support set for model training and the rest of data in each task as the query set for model evaluation. For B3, we implement three popular methods including **MAML** [4], **MatchingNet** [17], and **ProtoNet** [13] with cFeature for model optimization. Both MatchingNet and ProtoNet are metric learning based models and we use cosine distance for both models in this application. MatchingNet produces a weighted nearest neighbor classifier given the support set, while ProtoNet produces a linear classifier when cosine distance is used. MAML is a gradient-based model to learn well initialized model parameters which can be quickly adapted to new tasks and we set the base model of MAML as a 2-layer neural network. For B5, we feed the user embedding generated by six graph representation learning models mentioned above to MAML for drug traffickers detection.

3.2 Evaluation Metrics and Parameter Settings

To evaluate the performances of our model and baseline methods, we adopt two widely-used metrics: accuracy (**ACC**) and F1 score (**F1**). We apply Pytorch to implement all methods and all experiments are conducted under the environment of the Ubuntu 16.04 OS, plus Intel i9-9900k CPU, GeForce GTX 2080 Ti Graphics Cards, and 64 GB of RAM. For the meta-learning model, inner-level and outer-level learning rates are set as 0.05 and 0.08 respectively. Additionally, the optimal hyperparameters of ϵ, λ_{ssl}, and λ_{kd} are 0.95, 5, and 0.01 respectively. For graph representation learning models, the dimension of node embedding is 200, and the iteration number is 200. We run 50 times for each experiment by changing the value of random seeds and then we acquire the final average results.

3.3 Experimental Results

To validate the robustness and effectiveness of our model MetaHG, we conduct two sets of experiments on different tasks T1 (detecting opioids traffickers), and T2 (detecting hallucinogens traffickers) respectively. In Table 2, We can conclude that MetaHG is robust on both tasks and significantly outperforms other baseline models. Additionally, we find that R-GCN+MAML is the best baseline model, followed by HAN+MAML.

Table 2: Performance of accuracy$\pm 95\%$ confidence intervals of different models on each type/task (T_{id}).

Setting		5-shot		20-shot	
T_{id}	Model	ACC	F1	ACC	F1
T_1	GCN+MAML	0.7674 ± 0.012	0.7491 ± 0.008	0.7952 ± 0.009	0.7782 ± 0.006
	HAN+MAML	0.7947 ± 0.007	0.7749 ± 0.006	0.8374 ± 0.006	0.8253 ± 0.006
	R-GCN+MAML	0.8241 ± 0.005	0.8034 ± 0.005	0.8671 ± 0.005	0.8534 ± 0.004
	MetaHG	$\mathbf{0.8871} \pm 0.003$	$\mathbf{0.8442} \pm 0.002$	$\mathbf{0.9375} \pm 0.002$	$\mathbf{0.9324} \pm 0.002$
T_2	GCN+MAML	0.7634 ± 0.010	0.7457 ± 0.007	0.7885 ± 0.010	0.7741 ± 0.005
	HAN+MAML	0.7885 ± 0.006	0.7706 ± 0.005	0.8341 ± 0.005	0.8226 ± 0.003
	R-GCN+MAML	0.8185 ± 0.005	0.7991 ± 0.005	0.8617 ± 0.004	0.8515 ± 0.004
	MetaHG	$\mathbf{0.8825} \pm 0.003$	$\mathbf{0.8421} \pm 0.002$	$\mathbf{0.9321} \pm 0.003$	$\mathbf{0.9295} \pm 0.002$

4 Discussion and Limitation

Relied on the official Instagram APIs, all information we collected is public and we never collect any private user information or sensitive personal information. Additionally, all drug trafficker samples illustrated in this paper are anonymous and would not be harmful to the user. Therefore, we don't expect any privacy and ethical issues. Due to privacy regulations, the collected data will not be publicly accessible at this time while our model code is available at https://github.com/yyyqqq5/MetaHG. As collecting data on social media consumes extensive energy, we collect the data on Instagram as a showcase to analyze the drug traffickers on social media platforms. In the future, to effectively solve the drug trafficking problem, we wish to study these illegal activities on more social media platforms and further demonstrate the effectiveness of our model.

References

[1] Official instagram graph api, 2020. https://developers.facebook.com/docs/instagram-api.

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[3] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. metapath2vec: Scalable representation learning for heterogeneous networks. In *KDD*, 2017.

[4] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*, 2017.

[5] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.

[6] Jiawei Li, Qing Xu, Neal Shah, and Tim K Mackey. A machine learning approach for the detection and characterization of illicit drug dealers on instagram: model evaluation study. *Journal of Medical Internet Research*, 2019.

[7] Tim K Mackey and Janani Kalyanam. Detection of illicit online sales of fentanyls via twitter. *F1000Research*, 2017.

[8] NIH. National institute on drug abuse, 2021. https://www.drugabuse.gov/.

[9] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *KDD*, 2014.

[10] Sam Roweis. Em algorithms for pca and spca. *NeurIPS*, 1998.

[11] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *ESWC*, 2018.

[12] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[13] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *NeurIPS*, 2017.

[14] Donald F Specht et al. A general regression neural network. *IEEE Transactions on Neural Networks*, 1991.

[15] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S Yu, and Tianyi Wu. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *VLDB*, 2011.

[16] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[17] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. In *NeurIPS*, 2016.

[18] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. Heterogeneous graph attention network. In *WWW*, 2019.