Detecting Illegal Drug Usage in Social Media using Support Vector Machines and Deep Neural Networks

Introduction

Our objective in this project is to develop a system for identifying probable illegal drug users and distributors using social media. The primary structure of our approach can be separated very generally into two classification problems, one for identifying indications of drug use in text, and another to identify illegal drugs and drug paraphernalia in images. The broader impacts of this project are quite straightforward in the direct application of law enforcement. It would also provide potentially useful insight into central hubs of traffic through geo-tagged information. Within the field of computer science this project will require the creation of an tagged image database, which will have broader uses as an evaluation metric for various object recognition systems. Additionally any advancements made in object detection would be applicable to many diverse projects requiring such capabilities.

Background

The primary bases of this research are in text processing and object recognition. As such it will rely heavily on previous research in these fields for its necessary foundation. Our text processing classifier will be based on a single SVM classifier for identifying textually expressed intent to use or distribute illegal drugs based on the recent work by Wang et al. in intent mining on Twitter [1]. This technique has been selected over other more advanced methods due to the more informal nature of social media. This difference in structure creates difficulties for more recent methods such as the Stanford NLP parser [2]. In contrast to the bag-of-words model used by the SVM model in [1], we propose the use of the Word2Vec model for producing feature vectors [3]. This model has demonstrated significant performance gains over bag-of-words models over the past several years.

Our object recognition model will be based on recent work in object recognition using deep neural networks, as well as more established work using linear part-based methods [4] [5]. These techniques demonstrate considerable performance on a large variety of objects with the deep neural
network model demonstrating considerably higher accuracy in targets with high mutability. These also have the advantage of training in a more unsupervised manner, with the primary cost of increased incremental computational time after the model has been constructed. As both methods achieve state of the art performance, both are viable options for implementation in the image classifier of our drug detection model.

**Methods and Design**

This project will build upon existing systems in NLP and image classification in order to mold some newly developed ideas and techniques into a workable system for identifying drug use through social media. Initially this will require a preliminary filter for identifying images and text for further consideration. This rough filter will be followed by separate image and text classification systems. Final output from these classifiers given a sufficiently low false positive rate would be useful information for those in law enforcement or the drug enforcement agency.

Due to the nature of social networks, posted images are very frequently tagged with a text description. This description can be fed through the same preprocessing filter as independent text searching only for some drug related reference through a simple keyword search. Images with flagged descriptions, and independent text with matched keywords can then be fed separately into their respective classifiers.

The pure text classifier as previously noted will use an SVM model based on input vectors created from incoming text with a Word2Vec model. Text flagged for intent above a variable threshold will be identified for manual inspection. This is quite critical as it is imperative that text expressing clear intent be separated from innocuous reference. The variable threshold should counter this, allowing for a significantly reduced false positive rate creating a data output stream which is considerably more useful for practical applications despite the potentially decreased overall accuracy.

Unfortunately due to a lack of a proper image classification database related to the particular category of drugs and drug paraphernalia it will be necessary to construct such a training set in the course of this project. This will require either the manual identification of a large set of images containing certain objects, or a somewhat smaller sample of images in which the target objects have been tagged with bounding boxes for training and evaluation purposes. A large number of untagged images are available already on social media sites, and potential dataset images could be flagged by the preprocessing filter.
Once trained on this data the aforementioned object recognition models will be able to search images for objects in the database and place bounding boxes on any suspect portions of the image. These images along with their descriptions will be passed on for further manual inspection if a sufficiently low false positive rate is achieved, and volume does not surpass reasonable limitations. Alternatively user account metadata from flagged images may be used to track patterns on user profiles and flag only individuals with a significant number of flagged posts for manual investigation. This would in effect both reduce volume and decrease false positive rate, again at some cost to detection rate.

Resources

In order to complete this project a certain minimum level of equipment and personnel will be necessary. The most basic requirement will be a small number of personal workstations on which to create and evaluate the proposed preliminary filter and subsequent classification algorithms. Additionally access to some significant volume of text and image data from various social networks will be necessary both for creating a workable dataset for training models as well as for eventual real time testing. Finally the compilation of this dataset will require additional funding to crowd source the labeling of images with bounding boxes with appropriate levels of cross labeling to guarantee consistency.

Intellectual Merit

This project would provide further understanding of intent determination from text in the informal setting of social networks. It would also generate a completely new set of labeled training and testing data for evaluating the performance of current and future object recognition systems.

Broader Impact

The proposed system will provide immediate benefit to law enforcement and public service groups. Additionally it will promote the visibility of the field of study and promote interest in future work.
Works Cited


