

The Summer Undergraduate Research Fellowship (SURF) Symposium  
6 August 2015  
Purdue University, West Lafayette, Indiana, USA

## Classification and Visualization of Crime-Related Tweets

Ransen Niu<sup>1</sup>, Jiawei Zhang<sup>2</sup>, and David S. Ebert<sup>2</sup>  
Department of Computer Science<sup>1</sup>, Electrical and Computer Engineering<sup>2</sup>, Purdue University

### ABSTRACT

Millions of Twitter posts per day can provide an insight to law enforcement officials for improved situational awareness. In this paper, we propose a natural-language-processing (NLP) pipeline towards classification and visualization of crime-related tweets. The work is divided into two parts. First, we collect crime-related tweets by classification. Unlike written text, social media like Twitter includes substantial non-standard tokens or semantics. So we focus on exploring the underlying semantic features of crime-related tweets, including parts-of-speech properties and intention verbs. Then we use these features to train a classification model via Support Vector Machine. The second part is to utilize visual analytics approaches on collected tweets to analyze and explore crime incidents. We integrate the NLP pipeline with Social Media Analytics Reporting Toolkit (SMART) to improve the accuracy of crime-related tweets identification in SMART. This paper can also be utilized to improve crime prediction for law enforcement personnel.

### KEYWORDS

Data Mining, Social Media, Data Visualization