Week 9-3: Paper Summaries

CE-510 Seminar: Social Media Mining

Student Name: Jiaqi Guo NetID: JGR9647

■ Threat detection in tweets with trigger patterns and

In this paper, the author presents a novel method for detecting threatening messages on Twitter based on trigger keywords and contextual cues. The system was tested on multiple large collections of Dutch tweets. The experimental results show that our system can successfully analyze messages and recognize threatening content.

Their threat detection pipeline comprises a preprocessor, and two classifiers which act like a funnel. The first-stage classifier is basically a simple filter which weeds out the bulk of the incoming tweets. And their second stage classification is based on (1) mining positive and negative contextual cues for the triggers (2) Mine threat patterns from their training set of death threats using a sequence alignment technique which has its roots in the field of bioinformatics

Possible Improvement Directions:

1. Future work may include the development of a real-time early warning system for threats and a more effective combination of both approaches.

■ A Probabilistic Generative Model for Mining Cybercriminal Networks from Online social media

The main contribution of this paper is to develop a new weakly supervised network crime network mining method to facilitate network crime forensics. In particular, the proposed method is enhanced by a probabilistic generation model based on a new context-sensitive Gibbs sampling algorithm.

Their concept-based cybercriminal relationship classification method is more promising than keyword-based methods which relies on semantics rather than syntax. Concept-based approaches have already been shown to perform better than word-based methods for tasks such as topic modeling, domain adaptation and opinion mining.

Possible Improvement Directions:

1. Future work will likely include research into soft computing methods, such as genetic algorithms, to search for optimal or near-optimal system parameter values to further improve the effectiveness of the proposed approach. At the same time, in order to further evaluate the effectiveness and scalability of the proposed method, large-scale empirical experiments should be carried out to extract more online social media corpora.