Understanding Cyber Attack Behaviors with Sentiment Information on Social Media

Kai Shu[1], Amy Sliva[2], Justin Sampson[1] and Huan Liu[1]

[1] Computer Science and Engineering, Arizona State University, Tempe, AZ, USA
{kai.shu, jsampso1, huan.liu}@asu.edu
[2] Charles River Analytics, Cambridge, MA, USA
asliva@cra.com

Abstract. In today's increasingly connected world, cyber attacks have become a serious threat with detrimental effects on individuals, businesses, and broader society. Truly mitigating the negative impacts of these attacks requires a deeper understanding of malicious cyber activities and the capability of predicting these attacks before they occur. However, detecting the occurrence of cyber attacks is non-trivial due to the anonymity of cyber attacks and the ambiguity or unavailability of network data collected within organizations. Thus, we need to explore more nuanced auxiliary information that can provide improved predictive power and insight into the behavioral factors involved in planning and executing a cyber attack. Evidence suggests that public discourse in online sources, such as social media, is strongly correlated with the occurrence of real-world behavior; we believe this same premise can provide predictive indicators of cyber attacks. For example, extreme negative sentiments towards an organization may indicate a higher probability that it will be the target of a cyber attack. In this paper, we propose to use sentiment in social media as a sensor to better understand, detect, and predict cyber attacks. We develop an effective unsupervised sentiment predictor model utilizing emotional signals, such as emoticons or punctuation, common in social media communications, and a method for using this model as part of a logistic regression predictor to correlate changes in sentiment to the probability of an attack. Experiments on real-world social media data around well-known hacktivist attacks demonstrate the efficacy of the proposed sentiment model for cyber attack understanding and prediction.

Keywords: Cyber attack, sentiment analysis, social media

1 Introduction

As networked and computer technologies continue to pervade all aspects of our lives, the threat from cyber attacks has also increased. The broad range of increasingly common cyber-attacks, such as DDoS attacks, data breaches, and account hijacking, can have an extremely detrimental impact on individuals, businesses, and broader society. Thus, understanding these attacks and predicting

them before they occur is an emerging research area with widespread applications. However, detecting attacks, much less predicting them in advance, is a non-trivial task due to the anonymity of cyber attackers and the ambiguity of network data collected within an organization; often, by the time an attack pattern is recognized, the damage has already been done. Evidence suggests that the public discourse in external sources, such as news and social media, is often correlated with the occurrence of larger phenomena, such as election results or violent attacks. Social media, in particular, turns users into "social sensors" empowering them to participate in an online ecosystem that interacts with behavior in the physical world. We believe the same principle can apply to cyber attacks, where open source data may provide indicators to help understand the social and behavioral phenomena leading up to an attack.

In this paper, we propose an approach that uses sentiment polarity as a sensor to analyze the social behavior of users on social media as an indicator of cyber attack behavior. For example, extreme negative sentiment towards an organization may indicate a higher probability of it being the target of a cyber attack. However, measuring sentiment itself in social media is a challenging task due to: (1) the data challenge, where ground truth datasets with sentiment labels are often unavailable; and (2) the feature challenge, where effective and robust features must be extracted from short and noisy social media posts. Both challenges make standard supervised sentiment analysis methods inapplicable. Instead, we developed an unsupervised sentiment prediction method that utilizes emotional signals to enhance the sentiment signal from sparse textual indicators. In this model, we incorporate both emotion words and emoticons separately, as well as modeling the correlations among them in an unsupervised manner.

To explore the efficacy of sentiment polarity as an indicator of cyber-attacks, we performed experiments using real-world data from Twitter that corresponds to known attacks by a well-known hacker group. The experimental results show that the proposed sentiment prediction framework can recognize distinct behavioral patterns associated with these attacks. We also performed a temporal analysis on the sentiment for these attacks, which provides deeper understanding of the progression of ongoing cyber attack behaviors over time. Our contributions are summarized as follows:

a) We propose to utilize sentiment polarity in social media as a sensor to understand and predict the social behaviors related to cyber attacks;

b) We develop an unsupervised sentiment analysis using emotional signals, which models emotion indications without requiring labeled sentiment data beforehand; and

c) We conduct experiments on real-world Tweet data related to several cyber-attacks by a well-known hacker group to demonstrate the effectiveness of the proposed sentiment prediction framework.

2 Related Work

Our related work mainly falls into the following two categories: (1) sentiment analysis on social media; and (2) cyber attack analysis on social media.

Sentiment analysis on social media. Sentiment analysis has been an important task for natural language processing, and has been widely used in various social media applications, such as poll rating prediction [15], stock market prediction [2], fake news [21], emoji analysis [14] and so on. Existing methods can be categorized as either supervised [5, 8], meaning they are trained on labeled ground-truth data, and unsupervised [7, 9, 15], which are not trained on labeled data, but rather find patterns or groups in the existing datasets. Due to the lack of label information and the large-scale data produced by social media, unsupervised learning becomes more and more important in real-world social media sentiment analysis applications. Unsupervised methods often rely on a pre-defined sentiment lexicon to determine the sentiment score. The lexicon words are collected from (1) human annotators, such as in the General Inquirer [22] and Multi-Perspective Question Answering (MPQA) corpus [24] work; (2) a dictionary that contains semantic/linguistically related words, such as WordNet [16]; or (3) a corpus that can be used to infer sentiment polarity of words by exploring the relation between the words and some observed seed sentiment words in the corpus [15]. Recently, Hu *et al.* proposed a new state-of-the-art unsupervised sentiment analysis method that specifically leverages the way people communicate on social media, utilizing emoticon information, punctuation, and other sources of emotional signals to better predict sentiment on social media posts [7]. In this paper, we build on the success of this method to develop our sentiment predicting approach.

Cyber attack analysis on social media. In recent years, online social media has been a promising source of cyber attack analysis and understanding, such as threat intelligence fusion [13], malicious cyber discussion detection [12], etc. One line of research is to utilize social media platforms in specific domains to extract expert information as indication features [11, 18, 23]. In [11], Liao *et al.* utilize technology blog posts to extract key attack identifiers, such as source IP and MD5 hashes. Sabottke *et al.* estimate the level of interest in existing common vulnerabilities and exposures (CVE) and further predict the indication probability for real attacks [18]. Ritter *et al.* extract relevant Tweets of specific event with only a small set of supervised information to better collecting useful data for cyber decision making [17]. Recently, Kuandpur *et al.* used social media as a crowd-sourced sensor to gain insights into ongoing cyber attacks by adapting queries for searching Tweets, and better predict attacks, such as DDOS attacks [10].

3 Sentiment Sensor Modeling

In this section, we introduce our framework of predicting sentiment polarity in an unsupervised way. Then, we discuss how we build temporal sentiment analysis over time, which enables correlation of sentiment trends with other time series of real-world events, such as cyber attacks.

3.1 Unsupervised Sentiment Extraction

The proposed model is motivated by the observation that social media communication, such as Twitter, includes emotional signals (e.g., emoticons, specialized punctuation.) that could be strongly correlated with the sentiment in a social media post or the words in it. Our goal is to use the emoticons in social media posts to indicate the sentiment score of entire posts. Specifically, we aim to model the following emoticon information:

Table 1. List of Emoticons with Sentiment Polarity

Positive	**:-), (-:, =), (=, (:, :), :D, :d, d:, :), (:, 8), (8, 8), ;), ;), ;), (;, ;-), (-;, (;, ^ _ ^**
Negative	**:-(, :-(,)-:, =(,)=, :(,):, 8(,)8**

Post-level Emoticon Indication. Based on sentiment consistency theory [18], post level emotion indication assumes the strong correlation of sentiment polarity of a post and the corresponding emotion signals. The key idea of modeling post-level emotion indication is to make the sentiment polarity of a post as close as possible to the emotion indication.

Word-level Emoticon Indication: The overall sentiment of a post is also positively correlated with the sentiments of the words in that post. By modeling word-level emotional signals, we can utilize the valuable information in the sentiment analysis framework to infer sentiment polarity of a post.

To model post-level emoticon indication, we can build a classifier $y = f(\mathbf{x})$, where $y \in [0, 1]$ indicates the sentiment indicated by the emoticons themselves in the posts, and $\mathbf{x}$ represents the list of features can be extracted from social media posts. As shown in Table 1, there are commonly used emoticons that correspond to positive and negative sentiments. In addition, to model the word-level indication, we extract different types of features from post text, including platform-independent and platform-specific features, on social media.

For platform-independent features $\mathbf{x}^{(1)}$, we adopt the widely used n-gram features [4] with TF-IDF adaption to capture word-level patterns. We consider both the term frequency (TF) and inverse document frequency (IDF) to compute $\mathbf{x}^{(1)}$ in a post. Let $\mathcal{V} = \{w_1, w_2, ..., w_n\}$ denotes the vocabulary of the entire corpus, and $\mathcal{D} = \{d_1, d_2, ..., d_m\}$ denotes the set of all social media posts. Then the TF score for for word w_i in document d_j is computed as: $tf_{ij} = 1$ if $w_i \in d_j$, otherwise $tf_{ij} = 0$. The IDF measures whether a specific word is common or rare in the corpus, and it is computed as $idf_{(ij)} = \log \frac{m}{1+|d_j \in \mathcal{D}: w_i \in d_j|}$, where m is the total number of posts, $|d_j \in \mathcal{D} : w_i \in d_j|$ is the number of posts that word w_i appears in post d_j. Thus, we have the platform-independent feature vector computed as $\mathbf{x}_{ij}^{(1)} = tf_{ij} \times idf_{ij}$. The feature vector for each post is $\mathbf{x}_j^{(1)} = \mathbf{x}_{1j}^{(1)} \oplus \mathbf{x}_{2j}^{(1)} \cdots \oplus \mathbf{x}_{nj}^{(1)}$, and $\oplus$ is the concatenation operation.

For platform-dependent features $\mathbf{x}^{(2)}$, we aim to capture the specific linguistic patterns in the particular social media platform. In Twitter, for example, we apply the following heuristics to obtain additional features as shown in Table 2.

We selected these features because they provide useful indications of sentiment. For example, the question mark and exclamation marks can usually indicate a stronger sentiment strength. By concatenating all these features, we can obtain platform-dependent feature vectors $\mathbf{x}_j^{(2)}$ for post d_j. Finally, we combine platform-independent and platform-dependent features together, and get $\mathbf{x}_j = \mathbf{x}_j^{(1)} \oplus \mathbf{x}_j^{(2)}$.

We can next apply existing widely applied classifiers f to build a sentiment prediction model using these extracted features $\mathbf{x}$, such as Naïve Bayes, decision trees, logistic regression, K-nearest neighbors (KNN) clustering, and support vector machines (SVM). In this paper, we empirically adopt logistic regression as the classifier due to the fact that it is simple to train and understanding, but also very effective as a classifier. Note that even though the proposed model requires posts with emoticons to learn model parameters, it can predict the sentiment of posts without emoticons. The predicted sentiment score is represented by $\hat{y} \in [0, 1]$; the large the predicted value, the more positive the sentiment.

Table 2. Platform-specific features in social media posts

Feature	Description
HASHTAG	The number of hashtag in the post, e.g., #cybersecurity
QUESTIONMARK	The number of question marks (**?**) in the post
EXCLAIMATION	The number of exclamation marks (**!**) in the post
NEGATION	The number of negative words in the post, e.g., **not**.
TEXT_LEN	The length of the post by removing irrelevant mentions and URLs.

3.2 Temporal Sentiment Analysis

Sentiment time series have been widely used for event prediction, such as political election prediction [1], and disaster event detection [19]. Similar, in cyber attack scenario, we are not only interested in predicting the sentiment score of individual post, but also the temporal variation of sentiments over time. We aim to provide insights on: 1) characterizing how social media users change the sentiment polarity towards public events; 2) understanding how sentiments can indicate upcoming attack events; and 3) assessing the attack effects after the attack.

To tackle these questions, we collect the related social media posts $\mathcal{D}$ by querying relevant keywords (detailed in Section 4.1) in social media within a specific time range $\mathcal{T}$ that covers the time interval before, during, and after the attack event occurs. We build the sentiment time series $S = (\hat{y}_1, \hat{y}_2, ..., \hat{y}_m)$ chronologically by using the pre-learned sentiment predictor. We can also group and averaging the sentiments in different time granularities, such as per day, to explore the sentiment variations.

4 Experiments

In this section, we describe experiments we conducted using real-world datasets to demonstrate the effectiveness of the sentiment prediction approach described above and the ability to use this as a sensor for identifying cyber attack behavior over time and predicting future attacks.

4.1 Datasets

For our experiments, we used several different datasets. First, to empirically test the efficacy of using sentiment in social media as a sensor for indicating future cyber attacks, we first collected historical Twitter data using Gnip [3]. Gnip allows for historical queries against Twitter to be grouped by tags (i.e., topics) of interest. For looking at cyber attack behavior, we grouped our keywords into the following tags: (1) attack sources (i.e., tweets from or about known hacking organizations); (2) DDOS; (3) phishing; (4) exploits; (5) cyber security; (6) vulnerability announcements; (7) vulnerabilities; (8) CVEs; and (9) specific attack targets of interest. For each tweet, we used the sentiment analysis approach defined above to determine the sentiment value. We then aggregated over this data to create a time series with the mean sentiment scores per tag per day. It is used in combination with cyber incident reports provided by a financial company, C1, between April 2016 and September 2016, and a defense company, C2, between November 2016 and February 2017 [4] to develop a predictive model of cyber attacks (see Section 4.2). The reports provide details on three kinds of attacks: malicious email, malicious URL, and malware on endpoint.

In addition, to analyze the behavioral patterns associated with cyber attacks, we also collected another dataset from Twitter related to a well-known hacktivist group. We identified three attack events perpetrated by this group from 2016-2017, designated A_1, A_2, and A_3 [5]; these are different from the cyber attack incident data described above because these focus on hacktivist attacks that are known to be reactions to certain societal events, rather than typical cyber behavior targeting individual companies. Note that these attacks are sometimes benign behaviors, which means they are performed not for malicious intent (e.g., stealing credentials.) but more as a form of online protest or demonstration by a group seeking to influence societal events. Based on the three selected attacks, we developed specific keywords that were use to query GNIP, gathering tweets that occurred up to 3 weeks before and 1 week after the attack.

4.2 Experimental Results

Sentiment Clustering. Our first experiment seeks to evaluate the effectiveness of extracted features $\mathbf{x}$ for sentiment prediction, using cluster analysis to assess the performance of our unsupervised sentiment model. Because we do not have access to ground truth, we cannot compute standard accuracy measures. However, a cluster analysis will enable us to measure the quality of the patterns discovered by the unsupervised sentiment analysis. We try to answer the following questions: (1) Are the proposed sentiment features able to cluster all the tweets into distinct clusters that match intuitive understanding of sentiment? and 2) What is the proper cluster size we should use decide to sentiment degree?

[3] http://support.gnip.com/

[4] The names of the companies have been anonymized

[5] The attack events are anonymized here

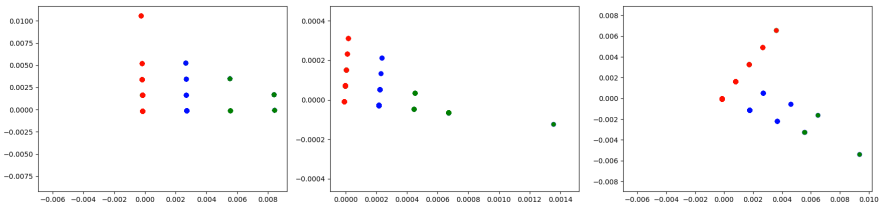
To answer these questions, we use k-means clustering [6] based on an extracted feature vector $\mathbf{x}_j$ for each post d_j, including the label assigned by our emotion-based sentiment analysis technique. The clustering performance is evaluated using the standard concepts of separation (i.e., the difference between elements in different clusters) and cohesion (i.e., the similarity of elements in the same cluster) captured in the widely used silhouette score metric. The Silhouette Score s is defined as $s = \frac{1}{m}\sum_{j=1}^{m}\frac{b(d_j)-a(d_j)}{\max(b(d_j),a(d_j))}$, where $a(d_j) = \frac{1}{|C|}\sum_{d_k\in\mathcal{D},d_k\neq d_j}\|\mathbf{x}_j-\mathbf{x}_k\|^2$ indicates the within-cluster average distance (cohesion) in cluster C, and $b(d_j) = \min\frac{1}{|G|}\sum_{d_k\in G}\|\mathbf{x}_k-\mathbf{x}_j\|^2$ indicates the distance of d_j with posts in other clusters (separation). Note that $s\in[-1,1]$, and the higher the score, the clusters show better separation from each other and a greater degree of internal consistency. If our unsupervised sentiment analysis approach is successful, it will produce results that have a high silhouette score across clusters that seem consistent with different sentiment categories. We also applied Priciple Component Analysis (PCA) for feature dimension reduction to better visualize the results of our cluster analysis. The results for A1, A2, and A3 are shown as in Table 3.

Table 3. Results of Post Clustering on Sentiment Features

Dataset	PCA	2		3		No	
	Cluster Size	2	3	2	3	2	3
A1	Silhouette Score	0.914	0.976	0.865	0.908	0.865	0.925
A2	Silhouette Score	0.921	0.981	0.803	0.914	0.853	0.902
A3	Silhouette Score	0.908	0.986	0.917	0.942	0.903	0.914

We can see better silhouette scores when we use three clusters over the sentiment feature space (i.e., positive, negative, and neutral) in all three cases. The high scores also indicate that the clusters are well separated and internally cohesive, indicating that the sentiment prediction model is able to use the emotional signal features to classify sentiment with a high degree of discrimination.

We apply PCA to project the original feature space to low dimensions for easy visualization. As shown in Figure 1, we can see that the cluster analysis results for the A_1, A_2, and A_3 attacks. In all cases, we observe three very distinct clusters for positive, negative, and neutral sentiment, which is consistent with the high silhouette scores.



(a) A1 (b) A2 (c) A3

Fig. 1. Cluster analysis visualizations case studies on hacker event dataset

Temporal Variation Analysis. We conducted temporal sentiment analysis on attack events A_1, A_2, and A_3 . The motivation to analyze sentiment variation is that it could provide informative behavioral indicators to predict attack events based on trends in public discourse on social media.

As shown in Figure 2 and Figure 3, the average sentiment scores are strongly correlated with the public event that preceded the attack. We have the following observations; (1) before the attack happens, the sentiment scores tend to be relatively stable, which may indicate the normal public discussion among users about a particular event; (2) while several days before the attack happens, the sentiment scores are very strongly negative, which may reflect the general public's unsatisfied attitude towards the event and indicate the potential for an upcoming attack; and (3) after the attack happens, the sentiment tends to increase again, which may indicate the positive response of social media users to the attacks (or the changes in the discussion precipitated by the attacks). Thus, there are distinct behavioral patterns in sentiment over time for indicating cyber attacks.

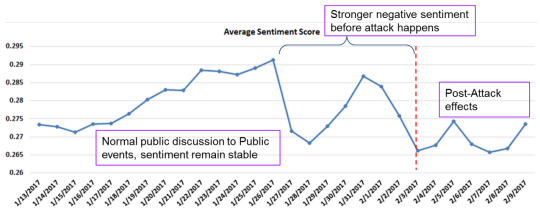


Fig. 2. Sentiment temporal variations on attack A1.

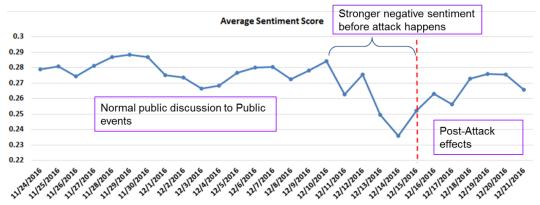


Fig. 3. Sentiment temporal variations on attack A2.

Cyber attack prediction using sentiment. Finally, we evaluate the sentiment trends for actually predicting cyber attacks before they occur. We use our historical data from Twitter and the cyber incident reports from C1 and C2 to develop a logistic regression classifier for each event type (i.e., malicious email, malicious URL, and endpoint malware). The features are the aggregate sentiment scores per tag/topic of interest, and the classes are the probability of an attack occurring. We varied the time lag between the sentiment scores and an attack between 0 and 10 days. We divide the data into a training set and a testing set in which the training set includes the first 80% of the time period the data covers, and the test set includes the other 20%. In each test, the logistic regression models predict whether or not an attack of each type occurs.

The results are summarized in Table 4. We found that the model performs much better in predicting either malicious-email or endpoint-malware attacks as

Table 4. Results for predicting endpoint malware attacks using sentiment sensor

Attack Type	Warning Threshold	Accuracy	Precision	Recall	F1
Malicious Email	*0.1*	0.5	0.905	0.5	0.644
	0.2	0.5	0.905	0.5	0.644
	0.3	0.905	0.905	0.5	0.580
	0.4	0.905	0.905	1	0.856
	0.5	0.905	0.905	0.905	1
Malicious URL	*0.1*	0.5	0.170	0.5	0.253
	0.2	0.5	0.170	0.5	0.253
	0.3	0.5	0.170	0.5	0.253
	0.4	0.5	0.170	0.5	0.253
	0.5	0.170	0.170	0.170	1
Endpoint Malware	*0.1*	0.5	0.801	0.5	0.615
	0.2	0.5	0.801	0.5	0.615
	0.3	0.5	0.801	0.5	0.555
	0.4	0.801	0.801	1	0.802
	0.5	0.801	0.801	0.801	1

opposed to the malicious-URL attack type, with very high precision and recall scores for both of these. Our data also showed that the time lag for Twitter events is rather small, with more successful prediction occurring with a time lag of between 1 and 3 days. In addition, more variation in the false positive and true positive rate is seen at even higher thresholds between 0.6 and 0.7, and using these we are able to generate the ROC curve as shown in Figure 4 for the malicious email event. This indicates that while sentiment shows promise as a predictor of cyber attacks, it is still only a weak signal and may need to be combined with other evidence or further amplified.

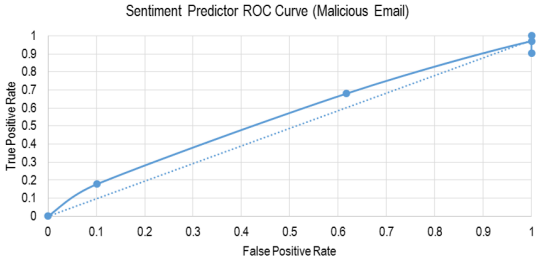


Fig. 4. The ROC curve for sentiment prediction for C2 malicious email attack

5 Conclusion and Future Work

In this paper, we use sentiment in social media as a sensor for understanding and predicting cyber attacks. The proposed sentiment extractor works in an unsupervised way utilizing emoticon signals for model learning. Experiments on real world datasets demonstrate the ability of sentiment score to (1) capture temporal correlations between attack events and inherent factors with ongoing public discourse; and (2) predict real-world cyber attacks, such as malicious email, malicious URLs, and endpoint malware against particular targets.

There are several interesting future directions. First, we can explore temporal process models, such as hawk process [3] for better modeling the sentiment variations over time for cyber attack prediction. Second, we can build temporal correlation networks [20] among general and specific attack Tweets to better predict the intensity of ongoing attacks. Third, we can explore other social features, such as credibility and veracity, to better understand the underlying social and behavioral patterns to help improve our cyber attack predictions.

Acknowledgements

This material is based upon work supported by, or in part by, the ONR grant N00014-16-1-2257 and N00014-17-1-2605, and the Office of the Director of National Intelligence (ODNI) and the Intelligence Advanced Research Projects Activity (IARPA) via the Air Force Research Laboratory (AFRL) contract number FA8750-16-C-0108. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, AFRL, ONR, or the U.S. Government.

References

1. Bermingham, A., Smeaton, A.: On using twitter to monitor political sentiment and predict election results. In: SAAIP'11
2. Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. Journal of computational science 2(1), 1–8 (2011)
3. Da Fonseca, J., Zaatour, R.: Hawkes process: Fast calibration, application to trade clustering, and diffusive limit. Journal of Futures Markets 34(6), 548–579 (2014)
4. Fürnkranz, J.: A study using n-gram features for text categorization. Austrian Research Institute for Artifical Intelligence 3(1998), 1–10 (1998)
5. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford 1(2009), 12 (2009)
6. Hartigan, J.A., Wong, M.A.: Algorithm as 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28(1), 100–108 (1979)
7. Hu, X., Tang, J., Gao, H., Liu, H.: Unsupervised sentiment analysis with emotional signals. In: WWW'13
8. Hu, X., Tang, L., Tang, J., Liu, H.: Exploiting social relations for sentiment analysis in microblogging. In: WSDM'13
9. IU, J.B., IU, H.M.: Twitter mood predicts the stock market (2011)
10. Khandpur, R.P., Ji, T., Jan, S., Wang, G., Lu, C.T., Ramakrishnan, N.: Crowdsourcing cybersecurity: Cyber attack detection using social media. arXiv preprint arXiv:1702.07745 (2017)
11. Liao, X., Yuan, K., Wang, X., Li, Z., Xing, L., Beyah, R.: Acing the ioc game: Toward automatic discovery and analysis of open-source cyber threat intelligence. In: SIGSAC'16
12. Lippmann, R.P., Campbell, J.P., Weller-Fahy, D.J., Mensch, A.C., Campbell, W.M.: Finding malicious cyber discussions in social media. Tech. rep., MIT Lincoln Laboratory Lexington United States (2016)
13. Modi, A., Sun, Z., Panwar, A., Khairnar, T., Zhao, Z., Doupé, A., Ahn, G.J., Black, P.: Towards automated threat intelligence fusion. In: ICIC'16

14. Morstatter, F., Shu, K., Wang, S., Liu, H.: Cross-platform emoji interpretation: Analysis, a solution, and applications. arXiv preprint arXiv:1709.04969 (2017)
15. O'Connor, B., Balasubramanyan, R., Routledge, B.R., Smith, N.A.: From tweets to polls: Linking text sentiment to public opinion time series. ICWSM'10
16. Peng, W., Park, D.H.: Generate adjective sentiment dictionary for social media sentiment analysis using constrained nonnegative matrix factorization. Urbana 51, 61801 (2004)
17. Ritter, A., Wright, E., Casey, W., Mitchell, T.: Weakly supervised extraction of computer security events from twitter. In: WWW'15
18. Sabottke, C., Suciu, O., Dumitras, T.: Vulnerability disclosure in the age of social media: Exploiting twitter for predicting real-world exploits. In: USENIX'15
19. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: WWW'10
20. Shu, K., Luo, P., Li, W., Yin, P., Tang, L.: Deal or deceit: detecting cheating in distribution channels. In: CIKM'14
21. Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H.: Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter 19(1), 22–36 (2017)
22. Stone, P.J., Dunphy, D.C., Smith, M.S.: The general inquirer: A computer approach to content analysis. (1966)
23. Tsai, F., Chan, K.: Detecting cyber security threats in weblogs using probabilistic models. Intelligence and Security Informatics pp. 46–57 (2007)
24. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: ACL'05