JES
Journal of Engineering Sciences

Vol 13, Issue 06, June /2022

ISSN NO:0377-9254

Detection of Possible Illicit Messages Using Natural Language Processing and Computer Vision on Twitter and Linked Websites

Mrs. A Veda Sri, Assistant Professor, Department of IT, vedasri.avirneni@gmail.com

V. Hemanth Reddy, BTech, Department of IT, hemanthveluri123@gmail.com

B. Pooja, BTech, Department of IT, poojabommireddy00@gmail.com

P. Sumiya, BTech, Department of IT, sumiyapathan111@gmail.com

Ch. Surekha, Department of IT, surekhachitrada2321@gmail.com

ABSTRACT: Human trafficking is a global problem that strips away the dignity of millions of victims. Currently, social networks are used to spread this crime through the online environment by using covert messages that serve to promote these illegal services. In this context, since law enforcement resources are limited, it is vital to automatically detect messages that may be related to this crime and could also serve as clues. In this paper, we identify Twitter messages that could promote these illegal services and exploit minors by using natural language processing. The images and the URLs found in suspicious messages were processed and classified by gender and age group, so it is possible to detect photographs of people under 14 years of age. The method that we used is as follows. First, tweets with hashtags related to minors are mined in real-time. These tweets are preprocessed to eliminate noise and misspelled words, and then the tweets are classified as suspicious or not. Moreover, geometric features of the face and torso are selected using Haar models. By applying Support Vector Machine (SVM) and Convolutional Neural Network (CNN), we are able to recognize gender and age group, taking into account torso information and its proportional relationship with the head, or even when the face details are blurred. As a result, using the SVM model with only torso features has a higher performance than CNN.

Keywords- *Support Vector Machine (SVM) and Convolutional Neural Network (CNN).*

1. INTRODUCTION

Initially the websites were isolated and just placed for reading since the user could not truly interact with the web. However, from the innovation and arrival of web 2.0, there was a revolutionary and radical change since the user stopped being a simple spectator and became an active individual in social networks such as Facebook, Twitter, Instagram, among others . Unfortunately, a door has also been opened for illegal businesses such as human trafficking , where some countries, such as Latin American countries, have the highest rates of smuggling of people, especially children and adolescents under 14 years old. It is important to note that the average age of consent is

14 years old in Latin American countries, so if underage people are used for illicit services areirectly considered victims of human trafficking. Currently, in Twitter, it is possible to find websites that offer escort or similar services where young girls are promoted for the consumption of ‘‘customers.’’ These girls are generally abused physically, psychologically, and sexually.

In recent years many criminal organizations advertise these ‘‘sexual services’’ using social networks hiding their illegal activity with seemingly innocuous terms such as ‘‘chicken soup’’ to refer to child pornography. Websites and social networks are used to extend this crime to the online environment, where covert advertising and messages are used to promote illegal services to exploit people who are victims of this crime, mainly minors. Although there are previous tweet filtering and image classification works to detect illicit messages, most of them use natural language processing methods or computer vision techniques separately. However, a different treatment of text and images is shown. In this paper, the authors focus their efforts on the analysis of advertisement published on the web for automatic detection of suspected messages. They use 10,000 ads manually annotated for this task. This work labels advertising that has text and images, and the analysis combines both types of information. They use a deep multimodal model called Human Trafficking Deep Network, and they obtained an F1 value of 75.3% with a recall of 70.9%.

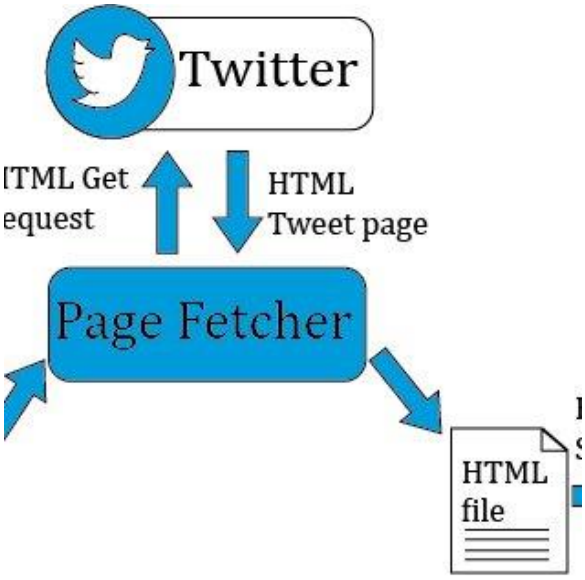

Fig.1: Example figure

On the other hand, the current image classification models use only facial information without taking into account that most of the images have the face blurred. In , the authors use computer vision algorithms to predict age with an approximate accuracy of 86.64%. In , SVM and CNN classification models are used to define the gender of a person. To the best of our knowledge, there are no works that consider characteristics of the upper body (upper torso) in the images to classify age groups.

2. LITERATURE REVIEW

2.1 A non-parametric learning approach to identify online human trafficking

Human trafficking is among the most challenging law enforcement problems which demands persistent fight against from all over the globe. In this study, we leverage readily available data from the website “Backpage”- used for classified advertisement- to discern potential patterns of human trafficking activities which manifest online and identify most

likely trafficking related advertisements. Due to the lack of ground truth, we rely on two human analysts-one human trafficking victim survivor and one from law enforcement, for hand-labeling the small portion of the crawled data. We then present a semi-supervised learning approach that is trained on the available labeled and unlabeled data and evaluated on unseen data with further verification of experts.

2.2 A new algorithm for age recognition from facial images

In this paper, a new algorithm for age-group recognition from frontal face image is presented. The algorithm classifies subjects into four different age categories in four key stages: Pre-processing, facial feature extraction by a novel geometric feature-based method, face feature analysis, and age classification. In order to apply the algorithm to the problem, a face image database focusing on people's age information is required. Because there were no such databases, we created a database for this purpose, which is called Iranian face database (IFDB). IFDB contains digital images of people from 1 to 85 years of age. After pre-processing, then primary features of the faces in the database will be accurately detected. Finally, a neural network is used to classify the face into age groups using computed facial feature ratios and wrinkle densities. Experimental results show that the algorithm identifies the age group with accuracy of 86.64%.

2.3 Detecting deception in text: A corpus-driven approach

Deception is a pervasive psycholinguistic phenomenon—from lies during legal trials to fabricated online reviews. Its identification has been studied for centuries—from the ancient Chinese method of spitting dry rice to the modern polygraph. The recent proliferation of deceptive online reviews has increased the need for automatic deception filtering systems. Although human performance is in general at chance, previous research suggests that the linguistic signals resulting from conscious deception are sufficient for building automatic systems capable of distinguishing deceptive documents from truthful ones. Our interest is in identifying the invariant traits of deception in text, and we argue that these encouraging results in automatic deception detection are mainly due to the side effects of corpus-specific features. This poses no harm to practical applications, but it does not foster a deeper investigation of deception. To demonstrate this and to allow researchers and practitioners to share results, we have developed the largest publicly available shared multidimensional deception corpus for online reviews, the BLT-C (Boulder Lies and Truths Corpus). In an attempt to overcome the inherent lack of ground truth, we have also developed a set of semi-automatic techniques to ensure corpus validity. This thesis shows that detecting deception using supervised machine learning methods is brittle. Experiments conducted using this corpus show that accuracy changes across different kinds of deception (e.g., lying vs. fabrication) and text content dimensions (e.g., sentiment), demonstrating the limitations of previous studies. Preliminary results confirm statistical separation between fabricated and truthful reviews (although not as large as in other studies), but we do not observe any separation between truths and lies, which suggests that lying is a much more difficult class of deception to identify than fabricated spam reviews.

2.4Use of technology in human trafficking networks and sexual exploitation: A cross-sectional multi-

country study

There are few data on the use of technology in human trafficking. This study attempted to address this gap in knowledge through field surveys that took place in India, Nepal, Thailand, Hungary, and the United Kingdom between 2010 and 2013. This research comprised face-to-face interviews with a total number of 246 individuals in 5 countries, consisting of 97 female victims, 64 traffickers, and 85 clients. The interviews were designed to help understand the role of technology such as the Internet, online social networking, and mobile phones in human trafficking. The survey was carried out using semi-structured questionnaires to find out how victims used technological devices under pre- and post-trafficking circumstances, how they advertised themselves, how diverse services and technologies were used to trade in sexually exploited trafficked people, and how clients explored, communicated, and paid for their sex transactions. The results showed that traffickers and their networks made good use of sophisticated software in order to safeguard their anonymity, make use of online storage and hosting services, and use advanced encryption techniques to counteract digital forensic investigations by the police.

2.5 A text-based deception detection model for cybercrime

Incidents of cybercrime exploiting text-based deception discourse are increasing due to popularity of text messages. We use machine learning and linguistic approaches to detect deception within text messages in cybercriminal networks. We develop cybercrime detection models by web genre. Our contributions are: models trained in scams in social media web genre detect fraud in messages in the email web genre with 60% predictive accuracy; models trained on fraud in the email genre can predict scams in social media web genre with 50% predictive accuracy. The prediction for the email model is promising due to the linguistic variations of cybercriminals in this study. We also demonstrate that cybercrime detection models can be constructed using features from natural language processing and linguistic psychological processes linked to cybercrime.

3. IMPLEMENTATION

Human trafficking is a global problem that strips away the dignity of millions of victims. Currently, social networks are used to spread this crime through the online environment by using covert messages that serve to promote these illegal services. In this context, since law enforcement resources are limited, it is vital to automatically detect messages that may be related to this crime and could also serve as clues.

DISADVANTAGES OF EXISTING SYSTEM:

Although there are previous tweet filtering and image classification works to detect illicit messages, most of them use natural language processing methods or computer vision techniques.

PROPOSED SYSTEM:

In this paper author is describing concept to detect human trafficking by analysing social media text messages with the help of SVM and Naïve Bayes machine learning algorithms. In this paper author first crawling twitter by using words like Lolita, escort and many more and then extracted tweets will go for cleaning to remove special symbols and stop words (words such as the, where, and, an, are etc.) and then tweets will be analyse to extracts words such as VERBS and ADJECTIVE and this words may

contains important subjects or suspicious words used by HUMAN TRAFFICKERS (the suspicious words can be chicken soup, girls, penguin and many more. Clean tweets will be given input to SVM and Naïve Bayes classifier to detect suspicious words.

If any tweet contains suspicious words then that tweet website will be scanned for images and each image will be processed through SVM HAARCASCADE classifier to detect face from that image and same algorithm will be used to detect upper body and both resultant images will be input to CNN (Convolution Neural Networks) classifier which will detect or predict AGE and GENDER from the resultant images. In this paper we are detecting gender as MALE and FEMALE and AGE will predicted with two classes as UNDER 14 Years or OVER 14 Years.

ADVANTAGES OF PROPOSED SYSTEM:

Using predictive models, such as Vector Support Machine (SVM) and Convolutional Neural Networks (CNN), the image classification process is done through a training phase and a testing phase.

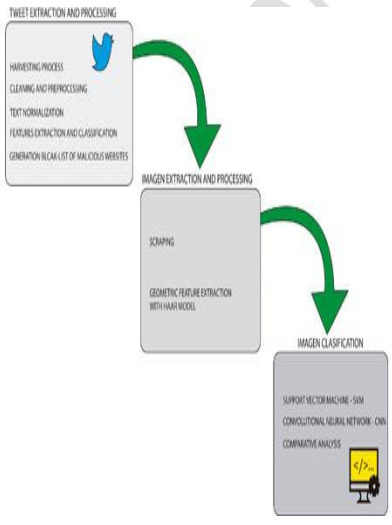

Fig.2: System architecture

The whole process for the detection of possible illicit messages related to human trafficking is presented. Firstly, the analysis and classification of tweets using natural language processing is done. Then, a blacklist of suspicious websites is obtained. Finally, the images related to the suspicious websites are processed and classified using Haar filters and SVM or CNN classifiers.

MODULES:

This project consists of following modules

1) Online Crawl Twitter: In this module we can enter HASHTAG and then application will crawl twitter using TWEEPY API to read all tweets from given hashtag.
2) Offline Upload Twitter Dataset: In this module if you don't want to crawl twitter then you can upload existing twitter dataset.
3) Clean Tweets & Extract Features: using this module each tweet will be processed to remove special symbols and stop words and then extract VERBS and ADJECTIVES and the clean tweets will be feed to SVM and Naïve Bayes algorithm. In both SVM and Naïve Bayes algorithms SVM is giving better suspicious tweets detection result.
4) Suspicious Tweets Classification using SVM & Naive Bayes: using this module we will input clean tweets to SVM and Naïve Bayes algorithms and then the application will divide entire data into train and test parts where 80% data will be used for training and 20% data will be used for testing. First by using 80% data algorithms will be trained and generate a model. A trained model will be applied on test data to

calculate prediction accuracy, precision, recall and FSCORE.

5) SVM & CNN Classification for Gender & age Prediction: After detecting suspicious tweets then each suspicious tweet website will be scan to read all images and then from that image face and upper body part will be extracted using SVM classifier and the resultant images will be input to CNN to predict AGE and GENDER.
6) Comparison Graph: in this module we are displaying comparison graph between SVM and Naïve Bayes in the form of precision, recall and FSCORE.

4. ALGORITHM

SVM:

Support Vector Machine(SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well its best suited for classification. The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. The dimension of the hyperplane depends upon the number of features. If the number of input features is two, then the hyperplane is just a line. If the number of input features is three, then the hyperplane becomes a 2-D plane. It becomes difficult to imagine when the number of features exceeds three.

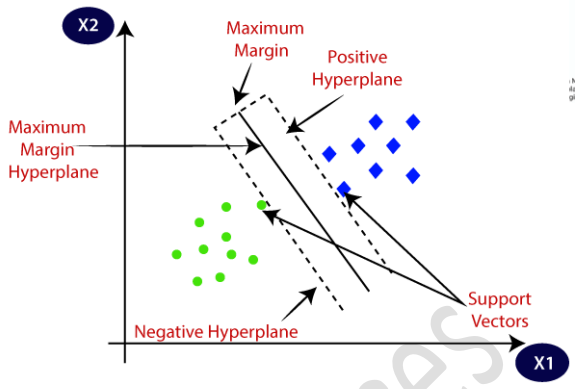

Fig.3: SVM model

The SVM kernel is a function that takes low dimensional input space and transforms it into higher-dimensional space, ie it converts not separable problem to separable problem. It is mostly useful in non-linear separation problems. Simply put the kernel, it does some extremely complex data transformations then finds out the process to separate the data based on the labels or outputs defined.

Advantages of SVM:

Effective in high dimensional cases

Its memory efficient as it uses a subset of training points in the decision function called support vectors

Different kernel functions can be specified for the decision functions and its possible to specify custom kernels

CNN:

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. CNNs are used for image classification and recognition because

of its high accuracy. It was proposed by computer scientist Yann LeCun in the late 90s, when he was inspired from the human visual perception of recognizing things. A Convolutional neural network (CNN) is a neural network that has one or more convolutional layers and are used mainly for image processing, classification, segmentation and also for other auto correlated data.

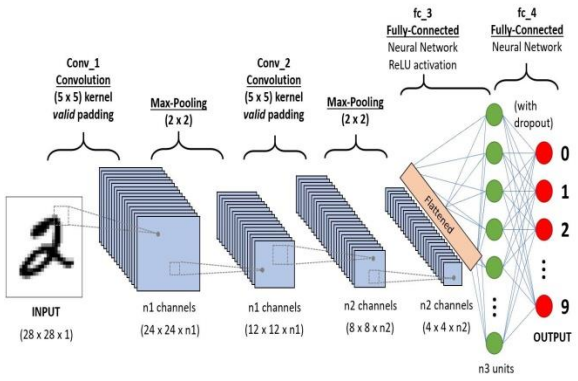

Fig.4: CNN model

Step by Step Guide

Step 1: Choose a Dataset. ...

Step 2: Prepare Dataset for Training. ...

Step 3: Create Training Data. ...

Step 4: Shuffle the Dataset. ...

Step 5: Assigning Labels and Features. ...

Step 6: Normalising X and converting labels to categorical data. ...

Step 7: Split X and Y for use in CNN.

5. EXPERIMENTAL RESULTS

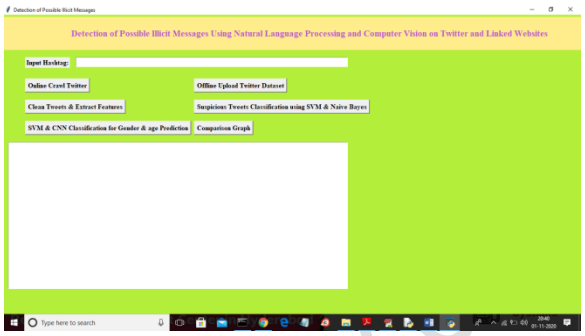

Fig.5: Home screen

In above screen I entered hashtag as ‘lolita’ and the press ‘Online Crawl Twitter’ button to start crawling. In below black screen we can see crawling started and I am displaying tweet date and tweet text

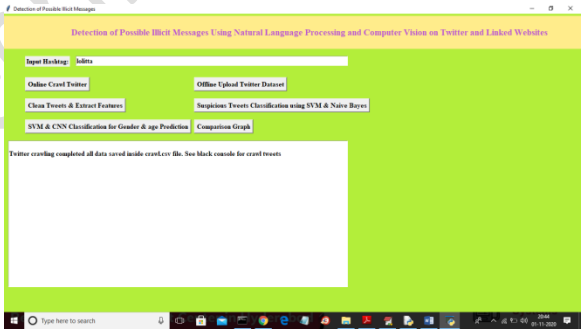

Fig.6: Clean tweets

In above screen we can see status message as twitter crawling complete and now click on ‘Clean Tweets & Extract Features

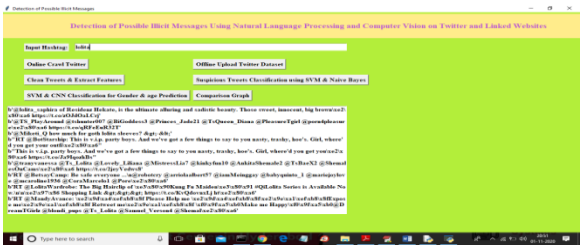

Fig.7: Features extracted

In above screen we can see each raw tweets that get processed for cleaning and now clean tweets are ready and to detect suspicious words click on 'Suspicious Tweets Classification using SVM & Naive Bayes' button to apply SVM and Naïve Bayes on each tweet to get suspicious words

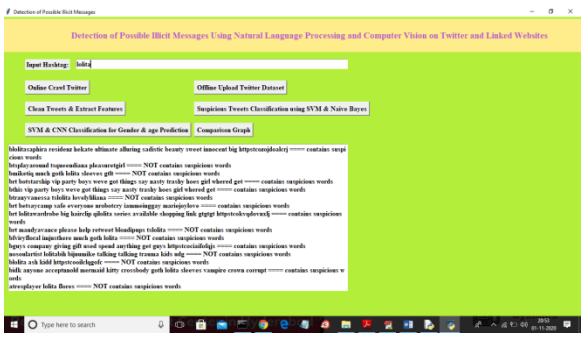

Fig.8: Suspicious Tweets Classification using SVM & Naive Bayes

In above screen displaying each cleaned tweets and after equal to symbol displaying detected result as contains suspicious words or not. Now we have tweets which contains suspicious words and now click on 'SVM & CNN Classification for Gender & age Prediction' button to scrape each tweets website to read image and the predict AGE and GENDER from images

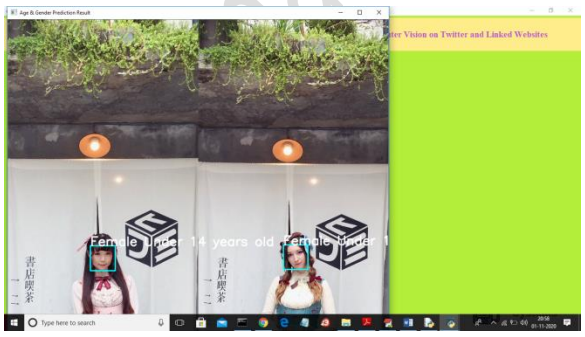

Fig.9: SVM & CNN Classification for Gender & age Prediction

In above screen application detected face and then displaying female under 14 years and application repeats above steps for all tweets.

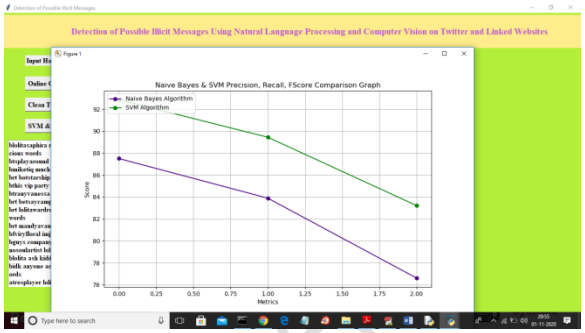

Fig.10: Comparison graph

In above graph blue line represents Naïve Bayes precision, recall and FScore and green line represents for SVM. In above graph x-axis contains precision, recall and FScore and y-axis represents its values. From above graph we can conclude that SVM is giving better performance.

6. CONCLUSION

Face recognition algorithms and machine learning models have been improved during the last years. For example, in the ILSVRC competition, an accuracy value of 90% +− 5% was obtained. In these conditions, machine learning recognition can be similar to visual object recognition used by human beings. Many factors have a direct impact on image recognition, such as size, color, opacity, resolution, kind of image format, among others. Therefore, the results of image recognition and classification depend on the dataset quality. In this work, we probed that satisfactory performance can be obtained using just geometric features of the torso and not only facial characteristics. For this paper, Haar filters combined with an SVM classifier were used for the extraction process of features, and then we classified the age

group and gender with an SVM classifier. The obtained results were compared with the outcomes of a CNN algorithm. SVM is a model widely accepted, and in this work, we obtained a classification accuracy higher than 80% for both experiments (face and upper body), not only for gender classification but also for age group classification. In this paper, our main contribution is the image classification based on the upper body to predict the age group to detect human trafficking. To the best of our knowledge, this work is the first approach related to image classification without facial features but just the upper-body geometric characteristics. Currently, there is no similar research that takes into account only the upper body features of minors. Thus, the results of this paper can be applied to human trafficking, disappearance, kidnapping, among others. Moreover, the obtained information can be used by the police or other security institutions.

7. FUTURE SCOPE

Finally, future work includes: 1) the study of some characteristics related to ethnic and racial features, 2) to extend the proposal to extract geometric features of the entire body, another kind of images, or inclusive videos in different formats, 3) detection of medical issues by means the analysis of features extracted from torso images, legs, back, among other characteristics, and 4) the use of other algorithms or the applicability in other networks like Instagram.

REFERENCES

[1] B. Bangerter, S. Talwar, R. Arefi, and K. Stewart, ‘‘Networks and devices for the 5G era,’’ IEEE Commun. Mag., vol. 52, no. 2, pp. 90–96, Feb. 2014.

[2] F. Laczko, ‘‘Data and research on human trafficking,’’ Int. Migration, vol. 43, nos. 1–2, pp. 5–16, Jan. 2005.

[3] M. Lee, ‘‘Human trafficking and border control in the global south,’’ in The Borders of Punishment: Migration, Citizenship, and Social Exclusion. Oxford, U.K.: Oxford Univ. Press, 2013, pp. 128–149.

[4] E. Cockbain and E. R. Kleemans, ‘‘Innovations in empirical research into human trafficking: Introduction to the special edition,’’ Crime, Law Social Change, vol. 72, no. 1, pp. 1–7, Jul. 2019.

[5] R. Weitzer, ‘‘Human trafficking and contemporary slavery,’’ Annu. Rev. Sociol., vol. 41, pp. 223–242, Aug. 2015.

[6] T. S. Portal. (2018). Twitter: Number of Monthly Active Users 2010-2018. [Online]. Available: https://www.statista.com

[7] M. R. Candes, ‘‘The victims of trafficking and violence protection act of 2000: Will it become the thirteenth amendment of the twenty-first century,’’ U. Miami Inter-Amer. L. Rev., vol. 32, p. 571, Jun. 2001.

[8] D. Hughes, Wilberforce Can be Free Again: Protecting Trafficking Victims. New York, NY, USA: National Review, 2008.

[9] A. Sultan, ‘‘Countering crime trafcking in persons smuggling migrants Ethiopia: The Law practice,’’ Ph.D. dissertation, School Law, Addis Ababa Univ., Ababa, Ethiopia, 2018, pp. 1–72.

[10] M. Tsikerdekis and S. Zeadally, ‘‘Online deception in social media,’’ Commun. ACM, vol. 57, no. 9, pp. 72–80, Sep. 2014.