

RESEARCH

Understanding Social Media Beyond Text: A Reliable Practice on Twitter

Qixuan Hou¹, Meng Han^{2*}, Feiyang Qu² and Jing (Selena) He²

*Correspondence:

menghan@kennesaw.edu

²Data-Driven Intelligence Research (DIR) Laboratory, Kennesaw State University, 30060 GA, USA

Full list of author information is available at the end of the article

Abstract

Social media provides high-volume and real-time data, which has been broadly used in diverse applications in sales, marketing, disaster management, health surveillance, etc. However, distinguishing between noises and reliable information can be challenging, since social media, a user-generated content system, has a great number of users who update massive information every second. The rich information is not only included in the short textual content but also embedded in the images and videos. In this paper, we introduce an effective and efficient framework for event detection with social media data. The framework integrates both textual and imagery content in the hope to fully utilize the information. The approach has been demonstrated to be more accurate than the text-only approach by removing 58 (66.7%) false-positive events. The precision of event detection is improved by 6.5%. Besides, based on our analysis, we also look into the content of these images to further explore the space of social media studies. Finally, the closely related text and image from social media offer us a valuable text-image mapping, which can enable knowledge transfer between two media types.

Keywords: social media; multimedia information; event detection

Introduction

The number of users on social media is huge. According to Our World in Data, there are 7.7 billion people in the world, with at least 3.5 billion of us online, which means one-in-three people in the world use social media [1]. Domo's Data Never Sleeps 5.0 report presents every minute of the day in 2017, there are 456,000 tweets sent on Twitter, there are 46,740 photos posted on Instagram, and there are 527,760 photos shared on Snapchat [2]. With such a large volume of real-time data, social media has become an important data source in both industry and academia. It is extensively used in a wide range of topics, including monitoring outdoor air pollution in London [3], modeling rumor spreading [4, 5], preventing sensitive information attacks, achieving disease surveillance [6], detecting natural disasters [7], etc.

With high-volume and real-time data, social media can, at times, outperform news sources on timeliness when reporting some types of events [8, 9]. For instance, it is quite a hot topic that the Amazon rain forest had alarming clusters of burning wildfires in the summer of 2019. There was a popular comment about the delayed media coverage, which stated that "When the Church of Notre Dame was burning, media coverage was all over the place in a few hours. The rain forest of Amazon in Brazil is burning for three weeks but no media coverage." Power from Media

Matters presented the bar chart in Fig. 1 to show the number of cable news segments mentioning the Amazon fires [10]. The first report on cable news was on August 21, 2019. However, the earliest tweet I can find about this wildfire is from August 6, 2019, as shown in Fig. 2.

Meanwhile, using social media data in the event detection system raises some challenges due to high volume, noises, and lack of geo-tagged information. When designing a social media-based application for landslide detection, we might face challenges as follows. As defined in Merriam Webster dictionary, “landslide” refers to “rapid downward movement of a mass of rock, earth, or artificial fill on a slope”, or “a great majority of votes for one side”. When collecting tweets with the keyword, “landslides”, we get tweets not only about natural disasters but also about elections, as shown in Fig. 3. Most social media limits the number of words in each post. For example, Twitter only allows 280 characters. Kokalitcheva shared that only about 1% tweets hit the 280-character limit, and 12% are longer than 140 characters [11], which means the textual content of each tweet can only carry limited information.

Several previous studies attempt to filter out noises and to understand the large data set extracted from social media. Harris *et al.* use geo-location information to remove irrelevant tweets [12]. Musaev *et al.* design text classification to filter out noises and compute the relevance ranking of users to achieve better accuracy on landslide detection [13, 7, 14]. McGough *et al.* include news sources, in addition to social media, to improve the accuracy on Zika incidence forecasting [15]. Researchers also explore topic analysis on social media data in the hope to extract meaningful topics from massive information. Kamath *et al.* and Argyrou *et al.* propose approaches to use hashtags as a major source to identify topics [16, 17] and Cataldi *et al.* and Han *et al.* use searching queries for topic detection [18, 19]. Most of the previous studies in event detection with social media focus on the textual content of social media.

More than likely, as shown in Fig. 3, imagery and textual content of one tweet are relevant, and images can carry high-quality and valuable information. With the rapid development of the internet speed and the promising future of 5G wireless networks, images and videos are becoming significant parts of social media. Image analysis is maturing as well. Machine learning and neural networks are available to automatically annotate images with keywords and to segment certain objects in the images. It’s time to fuse textual and visual aspects of data points and to further improve the performance of the existing social media-based applications.

In this paper, we investigate what benefits images can bring to social media-based event detection systems. We use Twitter data to detect landslides as a demonstration. We propose a three-stage pipeline. Three stages are Data processor, Detector, and Verifier. Data processor collects and pre-processes data. Detector is a dynamic candidate event management system, which can efficiently suggest and track candidate events. The candidate events will be evaluated and verified by Verifier. Verifier analyzes each candidate event from three perspectives, textual information, imagery information, and news. Our work proposes an efficient way to build image classification models for event detection on social media. We fine-tune a pre-trained CNN to classify images for the purpose of identifying landslides. Finally, we compare the detected events from the system with and without imagery content. It shows that

incorporating image analysis can help the system successfully reject 58 false-positive events out of 850 reported events. Besides, based on our analysis, we could also look into the content of these images which further explore the space of data analysis on social media. Our evaluation confirms that the image and text of one tweet are highly related. The closely related text and image offer us a valuable text-image mapping. Finally, we offer some insights into the valuable text-image mapping from social media. The mapping enables knowledge transfer between two media types. Unrestricted in social media studies, text-image mapping can also benefit a wide range of research.

The rest of the paper is organized as follows. Section 2 includes related works. Section 3 outlines the system overview. Section 4 focuses on discussing how Image verifier is built. Section 5 evaluates the proposed pipeline. Section 6 discusses other potential benefits that can be gained from imagery content. Section 7 concludes the paper.

Related work

A considerable amount of data on social media attracts researchers and companies from different fields. U.S. companies use social media, such as Twitter, to observe market trends and produce business values [20]. Twitter is used to track the spread of diseases and to monitor social commentary during the influenza H1N1 pandemic [21]. Social media technologies were deployed as the main knowledge sharing mechanisms among US government agencies during the 2010 Haitian earthquakes [22, 23].

Since users actively share real-life events on social media, event detection has been one of the major topics in social media studies. However, with large data volume and noises from social media, event detection is not an easy task. Previous studies propose different scalable approaches to filter out noises and identify events. We divide the previous studies into the following four categories.

Analyzing textual information. The textual content of social media has been extensively studied. Text classification is the process of assigning labels to text according to its content. It's one of the popular approaches used to analyze the textual content of social media. Musaev et al. design a landslide information system with Twitter data and use text classification to filter out irrelevant tweets to ensure the quality of detected events [7, 13]. Our proposed pipeline also includes a text classification to remove noises from our data set.

Cataldi et al. describe Twitter as “a low-level information news flashes portal” [18]. They create a navigable topic graph to present a set of emerging topics over time. Unlike our work trying to identify landslides, they don't have a target event type and, instead, they want to understand the latest hot topics.

Sentiment analysis is used to identify opinions expressed in the text, especially in order to understand if the attitude is positive, negative, or neutral. Yoo et al. propose Polaris, a system for analyzing users' sentimental trajectories for events analyzed in real-time [24]. The system can provide insights about events at a glance. Our future work will leverage sentimental polarity captured from textual content to better improve event detection accuracy.

Analyzing metadata. Metadata of social media include geo-location information, user profiles, hashtags, creation date and time, etc. Harris et al. propose a

Twitter-based food-borne illness reporting tool for the city of St. Louis [12]. Since they are only interested in the events in St. Louis, they set a boundary for tweet collection at a 50-mile radius around St. Louis using geographic coordinates as parameters. Geographic information helps them narrow down the number of relevant tweets. However, there are less than 1% tweets including geographic coordinates [25], indicating that there are tweets that might be relevant to food-borne illness in St. Louis but don't include geographic coordinates. Using geographic coordinates as parameters to collect tweets limits data set to at most 1% of tweets. In our work, we would like to monitor landslides worldwide. Therefore, instead of using geographic coordinates as parameters while collecting tweets, we collect data with keyword and apply Name-entity recognition (NER) on text to identify the location of the tweet.

A hashtag is defined as a word or a short term prefixed with the symbol “#”. It's widely used in social media including Twitter and Instagram and it's regarded as important metadata to categorize posts and to propagate ideas and topics. Kamath et al. and Argyrou et al. use hashtags as the main source for identifying topics from social media [16, 17]. Our future work will investigate how we can use hashtags related to our target event types to improve event detection efficiency and accuracy.

With creation date and time, we can study social media from the perspective of time series analysis. For instance, we can generate a time series by taking the total number of tweets collected from each day. Peak detection is to identify sudden surges, which might suggest the occurrence of an influential event. Healy et al. apply peak detection on social media data for event detection purposes [26]. We consider the creation date in our proposed pipeline, by including a sliding time window in Detector to dynamically manage candidate events.

User profiles are another important components of social media. Previous studies have proposed ways to identify influencers among social media users [27]. Influencers are individuals or organizations that have established credibility in a specific industry. Musaev et al. use Page Rank to identify influencers to improve landslide detection on Twitter data. They introduce the concepts of relevant and irrelevant virtual communities based on whether the posted messages are relevant to landslides or not and then apply the Page Rank algorithm to identify influential nodes in each community [13]. We would like to integrate user influence into our proposed pipeline in the future.

Analyzing external information. External information refers to data obtained from sources other than social media. To detect landslides, Musaev et al. combine social sensors and physical sensors. Social sensors are information collected from social media sources, including YouTube, Instagram, and Twitter. Physical sensors include seismometers for earthquakes and weather satellites for rainfalls [7]. McGough et al. include news sources and Zika-related Google search queries, in addition to social media data, to monitor Zika in five countries [15]. External information sources not only provide wider coverage than a single source but also improve the accuracy and reduce the latency overall. Our proposed pipeline also includes News verifier in an attempt to introduce another source of information to improve event detection accuracy.

Analyzing imagery information. With the rapid growth of internet speed, images are becoming a significant part of social media. Most of the existing event

detection approaches mainly focus on textual information and few studies explore imagery information and the correlation among the heterogeneous data. Alqhtani et al. propose to extract features from the text with the bag-of-words and from the image with histogram of oriented gradients (HOG) descriptors, grey-level co-occurrence matrix (GLCM), and color histogram. They fuse textual and imagery features and apply the K nearest neighbor classification to detect events [28, 29]. Papadopoulos et al. intend to automate the detection of landmarks and events. They create two image graphs representing two kinds of similarity between images, based on their visual features and their tags. The hybrid similarity image graph is generated by combining visual and tag features and the clustering algorithms are performed on the hybrid graph [30]. Both Alqhtani et al. and Papadopoulos et al. fuse imagery and textual features and machine learning algorithms are applied to combined features. In our work, we build two separate classification models for text and image and we can adjust the rules set up in Detector to decide whether to rely more on imagery or textual information. The approach provides flexibility to the pipeline. For different types of events, we can easily adjust the rules to rely more heavily on more relevant media types. Won et al. design a multi-task convolutional neural network for protest activity detection [31, 32]. They use Amazon Mechanical Turk (AMT) to obtain necessary annotations for each image, which implies that a lot of manual effort is involved in the process of building the model. In our work, we propose an automatic approach to prepare training image sets, which can dramatically decrease manual efforts. Our experiment results confirm the proposed approach prepares decent training data sets and the trained models produce promising classification results.

System overview

The pipeline of the social media-based event detection system is presented in Fig. 4. Three stages of the pipeline include Data processor, Detector, and Verifier. The pipeline is designed to collect social media data as input and detect events of interest as output. In our paper, we use Twitter to detect landslides for demonstration purposes. The pipeline can be easily configured to take other social media data sources (such as Instagram, Facebook, and Weibo data) or a combination of them to detect other event types (such as bridge collapse, highway breakdowns, road potholes, and California drought).

Data processor

Data processor, consisting of three major steps, first collects data from designated social media sources, then cleans data with stop words, and finally identifies location information.

Downloader

Multiple social media platforms provide APIs which enable programmatic access to large data sets. Facebook, Instagram, and Twitter are the three major social media platforms. Facebook is typically used to connect with friends and family. Instagram is designed to share photos and videos. People mainly post their highlights and follow influencers on Instagram. Twitter is considered as the platform to connect to

the world and follow real-time information. Since our studies are interested in natural disasters, we choose to use Twitter for demonstration. Twitter APIs allow us to access a large number of texts and images based on search terms. We collect tweets with images from 2018 about landslides with search terms including “Landslide”, “mudslides”, etc.

Cleaner

There are noises in tweets. For instance, we are interested in detecting landslides as natural disasters. However, “landslide” can also be used to describe an election in which the victor wins by an overwhelming margin. When retrieving tweets with the keyword “landslide”, some collected tweets describe elections, which are regarded as noises in our pipeline. Cleaner is the first simple step to remove obvious noises with stop words. More advanced mechanisms are implemented in Verifier to further remove noises later in the pipeline. There are duplicates in tweets. Retweeting, a convention in Twitter, means users post messages originally posted by others. Some of the tweets we collect have been retweeted more than 800 times, which indicates that, in our data sets, there are about 800 copies of the same tweet. We also remove duplicates to prevent unnecessary computational costs.

Geotagger

A detected event has to be defined by the location and the date. Without spatiotemporal features, the detected events won’t be useful. However, less than 1% of Twitter data contains geo-coordinates even if Twitter provides service to include user locations [25]. Geotagger intends to retrieve location names (geo-terms) from tweets and then encode the names into latitude and longitude (geo-codes). Named-entity recognition (NER) aims to identify and categorize named entities mentioned in unstructured text. SpaCy, an open-source software library, features fast statistical NER [33]. We use SpaCy to extract geo-terms in tweets. Google Maps APIs are used to convert geo-terms into geo-codes.

Detector

Detector aims to group tweets into candidate events by spatiotemporal features and Verifier will analyze the candidate events to decide their trustworthiness. A dynamic tracking system is designed to manage candidate events efficiently and intelligently, aiming to reduce calculation costs and to improve event coverage. A demonstration is shown in Fig. 5 and the database design is shown in Fig. 6.

Downloader retrieves and processes tweets from Twitter data pool every 24 hours. The system can be easily configured to run more often as needed. Firstly, tweets are saved into the database table *TWEETS*, with *TwitterID*, *TEXT*, *IMAGE*, *DATE*, and *RetweetCOUNT*. Cleaner drops tweets which contain stop words and removes duplicates. Geotagger extracts geo-terms, encode them into geo-codes, and save *GeoTERMS* and *GeoCODES* into table *TWEETS*. *EventID*, *TextVERIFIED*, *ImageVERIFIED* remain empty for newly added tweets.

Detector scans through newly added tweets and assigns them to the existing candidate events by matching their *GeoTERMS*. *EventID* in *TWEETS* table, and *NewTweetsCOUNT* and *NewImagesCOUNT* in *EVENT* table will be updated to

reflect the matches. For the tweets which cannot be matched to any existing candidate events, a new candidate event will be created in table *EVENT*, with a unique *EventID*.

In Text verifier and Image verifier steps, pre-trained models will classify newly added images and texts, and update *TweetVERIFIED* and *ImageVERIFIED* to relevant and irrelevant. Corresponding entries in *EVENTS* table will be updated to reflect the results of Verifier.

DateSTART in *EVENT* table is the date when the candidate event is created, or it is the date of the earliest tweets in the group. *DateEND* is the date of the latest tweets in the group. The two fields define the time range for each event. As shown in Fig. 7, a new event can be confirmed based on a set of rules. The rules can be easily adjusted as needed. When an event cannot be confirmed, we would like to know if the events are old. If so, the event will be archived to reduce future calculation.

Verifier

Three-step Verifier is designed to filter out noises in the collected data and to ensure that detected events are relevant to our target event types. Most social media-based event detection applications focus on analyzing textual information. With the rapid development of the internet speed, images and videos are becoming significant parts of social media. In our pipeline, we take advantage of additional information from images to further improve event detection accuracy and coverage. Additionally, News verifier retrieves relevant news from transitional news sources, such as the New York Times and CNN, to confirm the trustworthiness.

Text verifier

Text classification models are built to filter out irrelevant information. We manually label some tweets as training data and train text classification models. Eventually, the trained models can classify a tweet as relevant or irrelevant to landslides as natural disasters.

The first step towards training text classification models is to extract features from tweets. Tweets are free texts. We need to transform free texts into numerical representations in the form of a vector. This is known as word embedding. The most popular approach is the bag-of-words. A text is represented as the bag of its words, disregarding grammar and word order. Term frequency - Inverse document frequency (TF-IDF) and bigram are two modifications that enhance the bag-of-words. TF-IDF considers inverse document frequency. Bigram regards a sequence of two words as one element. Word2vec is another technique that uses a neural network model to learn word embedding. Word2vec represents one word with a unique vector.

The second step is to use machine learning algorithms to learn the association between the label and the numerical representation of texts. In our work, we compare the classification accuracy of logistic regression, support vector machines (SVM), random forest, and neural networks. Finally, with the pre-trained models, we are able to filter out irrelevant tweets.

Image verifier

Image classification model is built to filter out irrelevant information. There are a lot of deep learning models proposed to achieve image classification. The convolutional neural network (CNN), a class of deep learning neural networks, represents a huge breakthrough in this field. A CNN typically has convolutional layers, ReLU layers, pooling layers, and a fully connected layer. Convolutional layers apply a convolution operation to the input and pass the information to the next layer. The pooling layer combines the outputs of clusters of neurons into a single neuron in the next layer. Fully connected layers connect every neuron in one layer to every neuron in the next layer. Training a CNN requires a large amount of labeled images. With limited labeled images, we would like to fine-tune a pre-trained convolutional neural network. More details of how we prepare training images, how we build Image verifier and other design decisions will be introduced in the next section.

News verifier

Traditional news sources (such as New York Times, CNN, and CBS news) are widely regarded as reliable information sources, or we should state that compared to the reliability of news from social media, the reliability of traditional news sources is less questioned by the public. There are well-developed APIs, such as Bing News Search APIs, which provide quick and convenient access to local, national, and global news. We use the Bing API for demonstration purposes in our pipeline.

The news source is another verifier to further confirm the trustworthiness of detected events. With event type, location and, date, we query news APIs to search for news which potentially reports the same events. Events that can be found in news sources are typically significant events or the events that happen in major cities or well-developed countries and areas.

We won't reject the detected events which have no matched news. According to the previous studies, there are more landslides detected with information from social media than those reported by official landslide hazard reports from United States Geological Survey (USGS) [13]. Some non-influential local events are sometimes missed by news platforms and USGS. Furthermore, in undeveloped countries or areas, there are not enough news attention and official organizations to track events such as landslides. Our ultimate goal is to achieve event detection efficiently, accurately, and comprehensively. The pipeline aims to detect influential events as early as possible, preferably earlier than traditional news sources. Additionally, the pipeline can detect local and small events of interest with rich information from social media, especially the events in the areas which typically are not covered by traditional news sources. In the meantime, a three-step Verifier assures the reliability of detected events.

Collecting the events detected by social media and traditional news, we would like to study the differences between the two information sources. We are interested in the coverage, the trustworthiness, and the timeliness of event reporting by two sources.

Image verifier

With the rapid development of the internet speed, images and videos are becoming significant parts of social media, such as Instagram (images) and YouTube (videos).

Image analysis is maturing as well. Machine learning and neural networks are available to automatically annotate images and to segment certain objects in images. In our work, we won't focus on how to build the best image classification models. Instead, we want to present a promising way to automatically prepare annotated images and to train an image classification model that can be easily incorporated into existing social media-based event detection platforms to boost their event detection accuracy and coverage. Building an image classification model can be cumbersome for three reasons. Firstly, image classification is supervised learning, which means a large set of accurately labeled training data is required. The manual labeling process is tedious and time-consuming. Secondly, we need to decide how we'd like to classify images, or what categories we'd want to identify from a pool of images. Finally, training an image classification neural network typically requires a large number of training images. We'll address each of these difficulties in the following three subsections.

Image labeling

Text and image from one tweet are presumably related, which offers us a valuable text-image mapping. The mapping enables a knowledge transfer between the two types of information. Most of the previous social media-based event detection platforms focus on textual information and text filtering or text classification mechanisms have been built. With the existing mechanisms, we can label texts of tweets as relevant or irrelevant to the events of interest. Since we assume image and text from one tweet are related, we can label the image from the tweet as relevant or irrelevant based on the label of the text of the same tweet. Therefore, no manual image labeling is required anymore to prepare training image sets.

Image clustering

To understand the content of images and to decide how we'd like to classify images, we cluster images labeled as relevant and irrelevant separately.

Keras offers out-of-the-box with several CNN that have been pre-trained on the ImageNet data set. ImageNet project aims to manually label images into 22,000 categories for computer vision research. ImageNet Large Scale Visual Recognition challenge (ILSVRC) is an image classification challenge, intending to train a model that can accurately classify an image into 1,000 classes. Models are trained on about 1.2 million training images with another 50,000 images for validation and 100,000 images for testing. The state-of-the-art pre-trained networks suggest a strong ability to generalize to images outside the ImageNet data set via transfer learning, such as feature extraction and fine-tuning.

The VGG network (VGGNet) is a deep convolutional network developed and trained by Oxford's Visual Geometry Group (VGG), which achieved good performance on the ImageNet Challenge 2014 submission [34]. We use VGG16 for feature extraction purposes. The "16" stands for the number of weight layers in the network. The input layer takes an image of the size of $224 * 224 * 3$, and the output layer is a soft-max prediction on 1000 classes. The feature extraction part of the model is from the input layer to the last max-pooling layer, which is the size of 25,088 ($7 * 7 * 512$). We apply the K-means clustering method with 25,088 features to explore potential groupings.

Image classification

Since training an image classification network requires a large set of accurately labeled images, we would like to consider fine-tune a pre-trained network. Donahue et al. demonstrate that features extracted from a deep convolutional network, which is trained on a large and fixed set of object recognition tasks, can be reused for novel generic tasks [35]. It might be expected that the representations of a deep network are over-fitted for one particular task, as the network is discriminatively trained to perform well at one specific task. However, surprisingly, pre-trained networks often achieve better performance than that of hand-crafted features, especially when there are limited training images. The previous experiments find that early layers of CNN capture the general features of images, such as edges, lines, etc., and later layers capture more specific features, such as faces, shapes, etc. We can simply alter the later few layers or even only the output layer to achieve our classification purposes. Additionally, with limited training data set, training networks from scratch might lead to over-fitting. Pre-trained models are typically trained with larger data set and using the first few layers of pre-trained models help use reduce over-fitting. VGGNet is a publicly well known CNN for image classification and achieves good performance on the ImageNet Challenge 2014. VGGNet is quite large, with VGG16 as 533MB and VGG19 is 574MB. Chatfield et al. propose VGG-F network, a simpler version. It consists of 8 learnable layers, 5 of which are convolutional, and the last 3 are fully connected [36]. We take the existing VGG-F network, replace the final layer with random weights, and train the network again with images labeled as terrain or portrait. We can achieve 87% accuracy with five training epochs.

VGGNet is still used in some image classification problems, but smaller networks are designed after 2014 which can also achieve relatively good results with better efficiency, such as SqueezeNet [37] and Inception V3 [38]. In our future work, we would like to investigate other smaller pre-trained networks to further improve training efficiency without compromising classification accuracy.

In our work, we classify the text of tweets as relevant or irrelevant to landslides as natural disasters and then apply the same labels to the image of the same tweets. We cluster relevant and irrelevant images separately. Fig. 8 presents sample images from clustering results. For tweets relevant to landslides as natural disasters, we have three major clusters with images of map, terrain, and text. For irrelevant tweets, we have images of portraits, posters, and text. Clustering results identify two interesting image clusters, terrain and portrait. To help the landslide information system confirm or reject the detected events, we would like to build an image classification model that can identify terrains and portraits from a pool of images. We fine-tune a pre-trained VGG-F network to classify images as portrait or terrain. To validate its effectiveness and accuracy, we manually label 2000 images and image classification results are also carefully evaluated.

Experimental results show that the proposed way to automatically build Image verifier is promising. We can build Image verifier for any other social media-based event detection platforms efficiently without manual efforts. Introducing imagery information to event detection platforms might suggest better event coverage and accuracy, or, at least, images offer another angle of the detected events. For instance, Tien et al. propose the use of social media data to detect infrastructure breakdowns,

such as damage to bridges, highways, gas lines, and power infrastructures [39]. They focus on textual information from Twitter data. All Twitter data is run through a series of filters to obtain a subset of relevant data. With the proposed way to automatically build Image verifier, we can prepare training images without manual labeling; image clustering offers us insights into the content of images, and we fine-tune a pre-trained network. The image classification results might boost event detection accuracy and coverage, or it would be useful to provide images with detected events to show the severity of infrastructure damages.

Evaluation

To detect landslides, Downloader collects 438,000 tweets from 2018. Cleaner removes about 40%. Geotagger extracts geo-terms from 114,000 tweets. Since useful and valuable events should have spatiotemporal features, we only analyze those with geo-terms. Among these, 17,651 tweets, about 15%, have at least one image. Detector groups tweets by location and date. Candidate events are recorded in *EVENTS* table. We will evaluate the performance of the text classification model and image classification model separately. Finally, the evaluation of event detection accuracy with and without Image verifier will be reported.

We use the annotated landslides data set published by GRAIT-DM [40] to evaluate text classifiers. The data set contains about 4,000 tweets from 2014. The first three months' tweets are used as the testing set and the rest as the training set. We experiment with three types of word embedding techniques, TF-IDF, word2vec, and bigram and four types of classification techniques, logistic regression, SVM, random forest, and neural network. The different techniques produce similar classification accuracy. The detailed evaluation is shown in Table 1.

We use the trained text classifier to classify 2000 text of tweets with images from January 2018 as relevant or irrelevant to landslides and use the same label to annotate the images. Image clustering is applied and the sample clusters are presented in Fig. 8. As discussed in the previous section, we would like to build an image classifier to identify portraits and terrain from a pool of images. We annotate the images which are labeled as relevant to landslides and are clustered into a group of terrain as terrain and annotate those which are labeled as irrelevant and are clustered into a group of portraits as portrait. The remainings are labeled as other. We also manually label the 2000 images as portrait, terrain, and other. We fine-tune two image classification networks, one with manually labeled images, the other with automatically labeled images. The manual labels are used as ground truth for calculating image classification accuracy. The images are resized to 224 * 224 and normalized by subtracting the mean. With five training epochs, the model trained with manually labeled images can achieve about 86.2% accuracy and the model trained with automatically labeled images can achieve about 85.1% accuracy. Two models achieve similar results regarding the accuracy, which confirms our assumption that the image and text of one tweet are related and with a decent text classifier, to prepare image training data set, we can use the labels assigned to the text of tweets to annotate corresponding images. The detailed performance of the model trained by manually labeled images is shown in Fig. 9. The left pane shows the training loss and validation loss by training epochs. Whenever the network

makes mistakes, a loss is calculated, and the backpropagation algorithm updates the weights of the network in the direction that will decrease the loss. The middle one presents the training and validation accuracy with the top 1 error (how often the highest scoring label is wrong). The right one shows the top 5 error.

Finally, we compare the events detected by the pipeline with and without image classifier with tweets from February to December 2018. In order to simulate how Detector groups and manage candidate events dynamically, we process tweets day by day. We group tweets from February 1st by location and candidate events are generated and recorded in the database. We then move to Verifier stage and classify text and images to check if any candidate events can be confirmed. Table *EVENTS* and *TWEETS* will be updated accordingly before processing tweets from February 2nd. Our previous work does not include a dynamic event tracking system and it groups tweets by location and month [41]. The current Detector has a 28-day sliding window to group candidate events. As the window moves along the timeline, new tweets are added to the candidate events and old tweets are removed, which enables us to identify candidate events more efficiently and comprehensively.

Fig. 7 shows when to confirm or archive candidate events. The rules are calibrated based on our observation of tweets and candidate events from January 2018. Fig. 10 presents the summary of how Text verifier, Image verifier, and News verifier confirm and reject the candidate events. The proposed landslide detection pipeline reports 792 events in total. Among those, 161 events are confirmed by imagery content, 58 events are rejected by imagery content, and 60 events are confirmed by the news source. In order to show that imagery content can indeed boost event detection accuracy, we manually verify events rejected by Image verifier. All 58 rejected events are false events. Text verifier fails to correctly label the text of tweets, which lead to mistakenly confirm the candidate events. Table 2 presents the precision of event detection with and without images. Image verifier improves precision by 6.5%. The evaluation confirms Image verifier can help the system to reduce false-positive events and improve the accuracy of event detection.

Discussion

With the 4G enhanced mobile broadband capabilities, social media users dramatically increase and the content of social media is no longer limited to text. Now, 5G will further enhance the internet performance and we are experiencing the fast growth of embedded images and videos in social media. In the meantime, image analyzing techniques are maturing as well. Shifting from statistical methods to deep learning neural network methods, the field of computer vision has achieved state-of-art results on several interesting and practical problems. Deep convolutional neural networks can produce promising image classification results. Image classification with localization can not only assign a class label to an image but also show the location of the object in the image by a bounding box. Object segmentation splits an image into meaningful segments and object detection classifies a segment of an image. Technologies are also available to reconstruct an image by filling in missing or corrupt parts. Understanding images and extracting information from those can be easily achieved with all the advanced technologies available in the field of computer vision. It's time to fuse textual and visual aspects of social media and to study social multimedia.

In our proposed pipeline, we demonstrate an efficient approach for building an image classification model with training images labeled by text classification results. We compare image classification networks trained by manually labeled images and automatically labeled images. The accuracy of the two networks is similar. The results not only verify our proposed method but also show that the image and text of one tweet are closely related. The closely related text and image offer us valuable text-image mapping. The mapping enables knowledge transfer between two types of media. In addition to research related to social media, text-image mapping can also benefit a wide range of research that requires image or text data sets. For instance, we can find a lot of images of the Empire State Building from Twitter by collecting tweets with hashtag #empirestatebuilding. The text-image mapping can also initiate knowledge transfer from resource-rich media to resource-poor media. For example, sentiment analysis on textual data has made more progress than sentiment analysis on imagery data does, so the text is resource-rich media in the area of sentiment analysis and the image is resource-poor media. Thus, knowledge can be transferred from textual data to imagery data, making efforts to advance sentiment analysis with imagery data. Sentiment polarity of textual data can be leveraged to train sentiment classification models for imagery data.

Image verifier improves event detection accuracy by removing 58 false-positive events. The result reconfirms the benefits of having multimedia information, in offering different angles to analyze and validate the detected events. Especially for disaster response platforms, imagery content can be a valuable source for first responders to monitor the situations and identify the needs in the affected areas.

Conclusion

In this paper, We propose and evaluate a pipeline for event detection using social media data. For demonstration purposes, we use data from Twitter to detect landslide natural disasters. The proposed pipeline consists of three stages, Data processor, Detector, and Verifier. Detector is a dynamic event management system with a time sliding window to group tweets into candidate events and it helps the pipeline to identify candidate events more efficiently and comprehensively. Verifier contains three steps, Text verifier, Image verifier, and News verifier. Most previous social media event detection platforms only analyze textual data. Our work also studies imagery content. Additionally, we propose an automatic approach to efficiently build image classification networks for event detection purposes. The approach starts by automatically assigning text labels to images of the same tweets, clustering images to analyze their content, and finally fine-tune a pre-trained CNN to classify images into categories that we identify from the clustering results. The comprehensive evaluation proves the proposed approach is feasible and applicable, which implies that with the proposed approach, we can easily build image classification networks for other social media-based event detection platforms without manually labeling images and cumbersome training efforts. For instance, with the proposed approach, we can easily build an image classification network on top of the platform proposed by Tien et al. to detect infrastructure breakdowns [39]. Therefore, we can get both textual and imagery reports of infrastructure breakdowns. In addition, our work evaluates event detection with and without images and the result

indicates that incorporating imagery content appropriately can improve event detection accuracy. In our case, Image verifier successfully rejects false-positive events and the precision of event detection is improved by 6.5%.

In future work, we would like to evaluate our proposed pipeline and proposed an automatic way of building Image verifier on other event types with different social media sources, especially ones with more imagery content, such as Instagram. We will also use the proposed way to build Image verifier for the existing social media event detection platforms which currently only process textual information. Finally, many deep learning image classification networks have been developed and smaller network architectures have been proposed and evaluated in the past several years. We would like to experiment with other pre-trained networks, including SqueezeNet and GoogLeNet. A smaller and more efficient architecture is preferred in our work.

Authors' contributions

JH led the overall evaluation and guided the analysis structure. QH, MH and FQ participated in the design of the study and performed the text and image analysis. QH, MH, FQ, and JH co-authored the manuscript. All authors read and approved the final manuscript.

Availability of data

The data sets analyzed during the current study are available at <https://github.com/qixuanHou/SocialMediaBeyondText>.

Competing interests

The authors declare that they have no competing interests.

Funding

Funding information is not applicable.

Acknowledgment

We thank for all the valuable comments provided by the reviewers.

Author details

¹Georgia Institute of Technology, 30332 GA, USA. ²Data-Driven Intelligence Research (DIR) Laboratory, Kennesaw State University, 30060 GA, USA.

References

1. The Rise of Social Media. <https://ourworldindata.org/rise-of-social-media>
2. Data Never Sleeps 5.0. <https://www.domo.com/learn/data-never-sleeps-5>
3. Hsuen, Y., Qin, Q., Brownstein, J.S., Hawkins, J.B.: Feasibility of using social media to monitor outdoor air pollution in London, England. *Preventive Medicine* **121**, 86–93 (2019). doi:10.1016/j.ypmed.2019.02.005
4. Liu, L., Priestley, J.L., Zhou, Y., Ray, H.E., Han, M.: A2text-net: A novel deep neural network for sarcasm detection. In: *IEEE International Conference on Cognitive Machine Intelligence* (2019)
5. Han, M., Han, Q., Li, L., Li, J., Li, Y.: Maximising influence in sensed heterogeneous social network with privacy preservation. *International Journal of Sensor Networks* **28**(2), 69–79 (2018)
6. Nsoesie, E.O., Flor, L., Hawkins, J., Maharana, A., Skotnes, T., Marinho, F., Brownstein, J.S.: Social media as a sentinel for disease surveillance: What does sociodemographic status have to do with it? *PLoS Curr* (2016)
7. Musaev, A., Wang, D., Pu, C.: Litmus: A multi-service composition system for landslide detection. *IEEE Transactions on Services Computing* **8**(5), 715–726 (2015). doi:10.1109/TSC.2014.2376558
8. Han, M., Yan, M., Cai, Z., Li, Y.: An exploration of broader influence maximization in timeliness networks with opportunistic selection. *Journal of Network and Computer Applications* **63**, 39–49 (2016)
9. Albinali, H., Han, M., Wang, J., Gao, H., Li, Y.: The roles of social network mavens. In: *2016 12th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN)*, pp. 1–8 (2016). IEEE
10. The Notre Dame Fire Garnered Wall-to-wall Cable News Coverage. The Amazon Fires Are Barely Breaking Through. <https://www.mediamatters.org/msnbc/notre-dame-fire-garnered-wall-wall-cable-news-coverage-amazon-fires-are-barely-breaking>
11. A Year After Tweets Doubled in Size, Brevity Still Rules. <https://www.axios.com>
12. Harris, J.K., Hinyard, L., Beatty, K., Hawkins, J.B., Nsoesie, E.O., Mansour, R., Brownstein, J.S.: Evaluating the implementation of twitter-based foodborne illness reporting tool in the city of St. Louis department of health. *Int J Environ Res Public Health* (2018)
13. Musaev, A., Hou, Q.: Gathering high quality information on landslides from twitter by relevance ranking of users and tweets. In: *2016 IEEE 2nd International Conference on Collaboration and Internet Computing (CIC)*, pp. 276–284 (2016)
14. He, J.S., Han, M., Ji, S., Du, T., Li, Z.: Spreading social influence with both positive and negative opinions in online networks. *Big Data Mining and Analytics* **2**(2), 100–117 (2019)

15. McGough, S.F., Brownstein, J.S., Hawkins, J.B., Santillana, M.: Forecasting zika incidence in the 2016 latin america outbreak combining traditional disease surveillance with search, social media, and news report data. *PLoS Negl Trop Dis* (2017)
16. Kamath, K.Y., Caverlee, J., Lee, K., Cheng, Z.: Spatio-temporal dynamics of online memes: A study of geo-tagged tweets. In: Proceedings of the 22nd International Conference on World Wide Web. WWW '13, pp. 667–678. Association for Computing Machinery, New York, NY, USA (2013). doi:10.1145/2488388.2488447. <https://doi.org/10.1145/2488388.2488447>
17. Argyrou, A., Giannoulakis, S., Tsapatsoulis, N.: Topic modelling on instagram hashtags: An alternative way to automatic image annotation? In: 2018 13th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP), pp. 61–67 (2018)
18. Cataldi, M., Di Caro, L., Schifanella, C.: Emerging topic detection on twitter based on temporal and social terms evaluation. In: Proceedings of the Tenth International Workshop on Multimedia Data Mining. MDMKDD '10. Association for Computing Machinery, New York, NY, USA (2010). doi:10.1145/1814245.1814249. <https://doi.org/10.1145/1814245.1814249>
19. Han, S.C., Kang, B.H.: Identifying the relevance of social issues to a target. In: 2012 IEEE 19th International Conference on Web Services, pp. 666–667 (2012)
20. Culnan, M., McHugh, P., Zubillaga, J.: How large u.s.companies can use twitter and other social media to gain business value. In: *MIS Quarterly Executive* (2010)
21. Signorini, A., Segre, A.M., Polgreen, P.M.: The use of twitter to track levels of disease activity and public concern in the u.s. during the influenza a h1n1 pandemic. *PLOS ONE* **6**, 1–10 (2011). doi:10.1371/journal.pone.0019467
22. Yates, D., Paquette, S.: Emergency knowledge management and social media technologies: a case study of the 2010 haitian earthquake. In: *International Journal of Information Management*, vol. 31, pp. 6–14 (2011)
23. Gao, H., Barbier, G., Goolsby, R.: Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intelligent Systems* **26**(3), 10–14 (2011)
24. Yoo, S., Song, J., Jeong, O.: Social media contents based sentiment analysis and prediction system. *Expert Systems with Applications* **105**, 102–111 (2018). doi:10.1016/j.eswa.2018.03.055
25. Jurgens, D.: That's what friends are for: Inferring location in online social media platforms based on social relationships. In: Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media (2013). <https://www.aaai.org/Library/ICWSM/icwsm13contents.php>
26. Healy, P., Hunt, G., Kilroy, S., Lynn, T., Morrison, J.P., Venkatagiri, S.: Evaluation of peak detection algorithms for social media event detection. In: 2015 10th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP), pp. 1–9 (2015)
27. Han, M., Yan, M., Cai, Z., Li, Y., Cai, X., Yu, J.: Influence maximization by probing partial communities in dynamic online social networks. *Transactions on Emerging Telecommunications Technologies* **28**(4), 3054 (2017)
28. M. Alqhtani, S., Luo, S., Regan, B.: Fusing text and image for event detection in twitter. *The International journal of Multimedia Its Applications* **7**(1), 27–35 (2015). doi:10.5121/ijma.2015.7103
29. Han, M., Yan, M., Li, J., Ji, S., Li, Y.: Neighborhood-based uncertainty generation in social networks. *Journal of Combinatorial Optimization* **28**(3), 561–576 (2014)
30. Papadopoulos, S., Zigkolis, C., Kompatsiaris, Y., Vakali, A.: Cluster-based landmark and event detection for tagged photo collections. *IEEE MultiMedia* **18**(1), 52–63 (2011). doi:10.1109/MMUL.2010.68
31. Won, D., Steinert-Threlkeld, Z.C., Joo, J.: Protest activity detection and perceived violence estimation from social media images. *MM '17*, pp. 786–794. Association for Computing Machinery, New York, NY, USA (2017). doi:10.1145/3123266.3123282. <https://doi.org/10.1145/3123266.3123282>
32. Desai, S., Han, M.: Social media content analytics beyond the text: a case study of university branding in instagram. In: Proceedings of the 2019 ACM Southeast Conference, pp. 94–101 (2019)
33. spaCy. <https://github.com/explosion/spaCy>
34. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: *ICML* (2014). <https://arxiv.org/abs/1409.1556>
35. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: Decaf: A deep convolutional activation feature for generic visual recognition. In: *ICML*, pp. 647–655 (2014). <http://proceedings.mlr.press/v32/donahue14.html>
36. Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman, A.: Return of the devil in the details: delving deep into convolutional nets. In: *British Machine Vision Conference* (2014)
37. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K.: SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and 0.5MB model size (2016). 1602.07360
38. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the Inception Architecture for Computer Vision (2015). 1512.00567
39. Tien, I., Musae, A., Benas, D., Ghadi, A., Goodman, S., Pu, C.: Detection of damage and failure events of critical public infrastructure using social sensor big data, pp. 435–440 (2016). doi:10.5220/0005932104350440
40. The Annotate Landslide Dataset. <http://grait-dm.gatech.edu>
41. Hou, Q., Han, M.: Incorporating content beyond text: A high reliable twitter-based disaster information system. In: Tagarelli, A., Tong, H. (eds.) *Computational Data and Social Networks*, pp. 282–292. Springer, Cham (2019)

Figure 1 Cable news coverage on the Notre-Dame cathedral fire and on the Amazon rain forest fires. A bar chart prepared by Media Matters for America comparing cable news coverage on the Notre-Dame cathedral fire and the Amazon rain forest fires [10]

Figure 2 The tweet about the Amazon rain forest fires. An earliest tweet posted on August 6, 2019, about the Amazon rain forest fires.

Figure 3 The tweets with “landslides” keyword. Two tweets with “landslides” keyword. One is about natural disaster and the other is about election.

Figure 4 The infrastructure pipeline of social media-based event detection system. A three-stage event detection pipeline takes Twitter data as input and generates events with location and time range as output.

Figure 5 The candidate event tracking system. A demonstration of how Detector dynamically and efficiently group and track candidate events.

Figure 6 The database design. An example of two main tables, Tweets and Events, in the database.

Figure 7 The event life cycle. A decision tree showing how candidate event is confirmed or removed.

Figure 8 Three major categories of tweets relevant and irrelevant to landslides as natural disasters. Sample images from three major categories which are relevant and irrelevant to landslides as natural disasters.

Figure 9 The performance summary of image classification result. The training loss and validation loss, top 1 error and top 5 error by training epochs are displayed.

Figure 10 The summary of detected event results. The candidate events are verified by Text verifier, Image verifier, and News verifier.

Tables

Table 1 Evaluation of text classifier with the annotated tweets from 2014.

accuracy	TF-IDF	word2vec	bigram
logistic regression	0.89	0.87	0.90
SVM	0.89	0.89	0.88
random forest	0.89	0.89	0.88
neural network	0.89	0.88	0.87

Table 2 Precision of event detection with and without imagery content.

	# detected events	precision
text-only	850	89.8%
with images	792	96.3%