Geo-spatial Multimedia Sentiment Analysis in Disasters

Abdullah Alfarrarjeh, Sumeet Agrawal, Seon Ho Kim, Cyrus Shahabi
Integrated Media Systems Center, University of Southern California, Los Angeles, USA
{alfarrar,sumeetag,seonkim,shahabi}@usc.edu

***Abstract*— Sentiment analysis of disaster-related posts in social media can contribute to the situation awareness and better understanding of the dynamics of disaster events by identifying the polarity of sentiments expressed by the public. However, Even though many sentiment analysis techniques have been developed and available, there are still limitations in reliably using sentiment analysis since there is no dominantly accepted technique in disasters. Taking advantage of existing state-of-the-art sentiment classifiers, this paper proposes a novel framework for geo-spatial sentiment analysis of disaster-related social media data objects. Our framework addresses three types of challenges: the inaccuracy and discrepancy associated with various text and image sentiment classifiers, the geo-sentiment discrepancy among data objects in a local geographical area, and observing diverse sentiment from multimedia data objects (i.e., text and image). The extracted sentiments are aggregated geographically for the purpose of extracting more accurate local regional insights. For the evaluation of the framework, we explored Twitter and Flickr datasets at the time of Hurricane Sandy and Napa Earthquake and showed how our approach can provide a better understanding of disaster events.**

Keywords – sentiment analysis, disaster, geo-sentiment analysis, collective sentiment analysis, text sentiment analysis, visual sentiment analysis

I. INTRODUCTION

Online social media (e.g., Twitter, Flickr, Instagram, and Facebook) have attracted millions of users to communicate and share thoughts and feelings about their daily lives. Especially, people can share geo-tagged data objects (e.g., images or text) anytime and anywhere, which potentially provides a valuable source of knowledge about the physical environment and social phenomena. Many studies have utilized social media data to investigate human behaviors and public characteristics (e.g., demographic and urban characteristics [16] [8], customer opinions about products [14] [19], political views [17], and public health information [18]). Furthermore, social media has been used as active sensors during emergency events [15] [20] such as disasters (e.g., flood, earthquake, and hurricane). In particular, the Federal Emergency Management Agency (FEMA) identifies social media as an essential component of future disaster management [35].

Extracting people opinions and emotions from the textual and visual data objects posted in social media has been performed through a data mining technique known as sentiment analysis (a.k.a. sentiment classification). Sentiment analysis has been widely used and applied in various studies such as Oscar awardee prediction [25], presidential election [22], new product marketing [2], etc. Among many potential applications of sentiment analysis, this study focuses on disaster situations, especially combined with geographical properties which are naturally tied with disasters. Given a disaster situation, geo-tagged social media data objects contain rich information about the disaster (e.g., people's concerns and panics, warning messages, and devastation). Thus, sentiment analysis of disaster-related posts in social media provides a semantic abstract of to the situation awareness and better understanding of the dynamics of the disasters. Furthermore, spatial and temporal consideration of sentiment provides better local regional insights in making decisions at crisis management. Examples of research work related to sentiment analysis on social media in disaster include the studies performed by Chien et al. [46], who evaluated sentiment analysis of Flickr data in disaster management at the time of the strike of the typhoon in Taiwan 2009, and Dewan et al. [33], who analyzed the sentiment of textual and visual content obtained from Facebook during the terror attacks in Paris 2005.

Even though many sentiment analysis techniques have been developed and available, there are still issues in sentiment analysis. For example, most approaches consider only one type of data, mainly text, and provide proprietary sentiment labels resulting in a non-standard inconsistent representation of sentiments across different approaches. To understand the aggregated sentiment of people in a geographical area during disasters, it is required to have a new geo-sentiment model which is different from conventional sentiment models. While sentiment analysis has been extensively studied for a single data type such as text or image [46] [48], geo-sentiment analysis of a geo-tagged multimedia data objects obtained from a plethora of social media comprises several challenges, especially in disasters. First, the certainty of a data object's sentiment is correlated with its geographical neighbors affected by the same disaster. For example, an object whose sentiment contradicts those of its geo-neighbors may influence the overall sentiment certainty of the local geographical area. Second, there are various sentiment classifiers which inevitably suffer from inherent classification inaccuracy. Analyzing an object with different classifiers and resulting with various sentiment values decreases the certainty of the observed sentiment value of the object. Third, social media provides heterogeneous types of objects (e.g., text and image) which depict the disaster in different ways. Different object types have different values in terms of expressing the sentiment hence the importance of the each type varies. In particular, Tweets are short messages and contain informal words which might affect the accuracy of their corresponding classifiers. In addition, the information included in the text of a Twitter post may not be related to the actual geo-tagged location associated with the post.

In this paper, we propose a geo-sentiment model which addresses all of the above challenges using a comprehensive framework to analyze the geo-spatial sentiment from a heterogeneous set of social media data. Our framework is composed of three phases. In the first phase, we utilize multiple libraries of popular sentiment classifiers to label each object independently. Since, in disaster situations, the objects posted by people on social media show high binding to each other in terms of time and location, we partition geo-tagged objects temporally and then spatially in the second phase. In the final phase, we estimate the aggregated sentiment based on our proposed geo-sentiment model where the sentiment label of each object comprises sentiment scores generated by multiple classifiers in addition to the sentiment labels of its geo-neighbors. The rationale of such a sophisticated framework is to provide a collective and normalized sentiment from multiple data types using various analysis techniques, which subsequently enhances the understanding of disaster situation in a clearer way.

To evaluate our framework, we analyzed two cases of different disasters; Hurricane Sandy and Napa earthquake. First, we collected a heterogeneous dataset from Twitter and Flickr during the disaster time. After that, we analyzed the datasets and compared our results with FEMA reports about Sandy [41] and Napa [42]. Our framework was able to detect the disaster situation, and the visualization results were in correspondence with both FEMA reports. In the case of Napa earthquake, we also compared our results with the shakemap [43] provided by the United States Geological Survey (USGS).

The remainder of this paper is organized as follows. Section 2 introduces a set of preliminary definitions, background about the state-of-the-art techniques in sentiment analysis, and the problem statement. Section 3 explains our framework for evaluating sentiment. Section 4 reports on our experimental results. In Section 5, we review the related work. Finally, in Section 6, we conclude and discuss future work.

II. Preliminaries

A. Sentiment Analysis

Text sentiment analysis: Many researchers have devised various algorithms to analyze textual material for predicting sentiment. For example, Socher et al. [34] proposed a recursive neural network classifier referred to as CoreNLP. CoreNLP is based on the natural language pipeline that contains several steps including tokenization (i.e., chopping text into words, numbers, spaces, and punctuations), part of speech tagger (i.e., assigning a syntactic category to each token), and morphological analysis (i.e., finding the root form of each word). Another well-known sentiment classifier is NLTK [13], the Natural Language Toolkit for Python, which uses the Porter stemming algorithm (a simplified technique for morphological analysis) and is based on the naive Bayes classifier. Furthermore, Thelwall et al. proposed SentiStrength [10] which extracts the intensity of text by estimating the strength of positive and negative of each token in the text. SentiStrength showed good results for social web texts (i.e., short text). On the other hand, several commercial text sentiment classifiers are recently offered by the industry (e.g., Microsoft Azure Text Analytics API [28], Google Cloud Natural language API [1], and IBM Watson Alchemy Language API [12]).

Visual Sentiment Analysis: Recently, researchers have investigated extracting sentiment from visual content because visual data have become an essential part of social media. Current approaches of sentiment analysis on images are categorized into three mechanisms: low-level visual feature based methods [23] [9], mid-level visual feature based methods [24] [7], and deep learning based approaches (i.e., the convolutional neural network) [26] [27] [11]. Both Jia et al. [23] and Yang et al. [9] employed color features for the visual sentiment prediction. Borth et al. [24] proposed a visual sentiment ontology (referred to as SentiBank) composed of 1200 adjective-noun pairs. Similar to SentiBank, Yuan et al. [7] employed 102 scene attributes instead. Among the research efforts which utilized the convolutional neural network (CNN), Campos et al. [11] applied transfer learning to fine-tune the neural network with Flickr Dataset. In the industry, there are commercial products for facial expression prediction (e.g., Microsoft Azure Computer Vision API [5] and Google Cloud Vision API [6]) but these products are not able to predict the sentiment of “free of faces” images. Alternatively, these products offer other functions for extracting keywords describing image’s content. Thus, text sentiment classifier can be employed on the combination of the extracted keywords.

In general, both text and visual sentiment classifiers produce a sentiment label (i.e., s) as negative, neutral, or positive which corresponds to the numeric sentiment score -1, 0, and 1, respectively.

B. Problem Definition

DEFINITION 1 (Geo-tagged Social Media Data Object): Each object O is represented by four attributes: content $O.c$, timestamp $O.s$, geo-location $O.l$ (longitude and latitude), and type $O.p$ which can be an image, video, text or composite object (e.g., an image tagged with text).

DEFINITION 2 (Heterogeneous Geo-tagged Social Dataset): We have a dataset of n geo-tagged data objects $D = \{O_1, O_2, ..., O_n\}$ retrieved from various social media which provides different types of objects.

In this paper, we consider analyzing two independent datasets which include text objects (i.e., tweets) obtained from Twitter and image objects obtained from Flickr captured during the same period and in the same geographical area. In the presence of such datasets, we identify the following key challenges in geo-spatial multimedia sentiment analysis:

a) Classification Discrepancy: Sentiment analysis techniques are not yet mature, so even the state-of-the-art sentiment classifiers inevitably bring in classification inaccuracy. Moreover, there is no uniform evaluation metric to represent the results from different techniques. Therefore, when an object is analyzed using different sentiment classifiers, these classifiers potentially generate different sentiment labels.

b) Geo-sentiment Discrepancy: Geographical sentiment analysis in disaster situations aims at recognizing the areas where the majority of people interact with the same sentiment mood. Nonetheless, when analyzing a local region, it is possible to find that the sentiments of objects posted by individuals in the same area vary, which indicates geographical sentiment discrepancy.

c) Dataset Heterogeneity: Different social media provide various types of objects (e.g., texts in Twitter, images in Flickr) which may differently depict the situation happening in the geographic area. Moreover, a tweet does not necessarily reflect the sentiment of the sender's location while an image posted in Flickr certainly reflects the sentiment of its location.

Addressing the first two types of challenges requires estimating the certainty of object sentiment. For this sake, we propose a couple of sentiment aggregation approaches and define two types of committees per object as follows:

DEFINITION 3 (Classifiers Committee per Object): Each O is labeled by m sentiment classifiers (referred to as classifiers committee $\mathbb{C}_C$) based on $O.c$. Choosing sentiment classifiers for O depends on $O.p$.

Through $\mathbb{C}_C$ (O_i) consisted of m classifiers, O_i is associated with a set of sentiment labels $\{s_1, s_2,\ldots, s_m\}$. The certainty of an object sentiment is estimated by measuring the sentiment disagreement among the elements of $\mathbb{C}_C$.

DEFINITION 4 (Geo-neighbors Committee per Object): For a given spatial distance metric Φ and distance threshold σ, the geo-neighbors committee $\mathbb{C}_N$ of $O_i \in D = \{\exists O_j \in D \mid \Phi(O_i.l, O_j.l) \leq \sigma\}$.

Without loss of generality, we used the Euclidean distance as the Φ metric. Through $\mathbb{C}_N$ (O_i) consisted of k objects, each $O \in \mathbb{C}_N$ is associated with a sentiment label s (i.e., $\mathbb{C}_N$ (O_i) = $\{s_1, s_2,\ldots, s_k\}$). The certainty of an object sentiment is estimated by measuring the sentiment disagreement among the elements of $\mathbb{C}_N$.

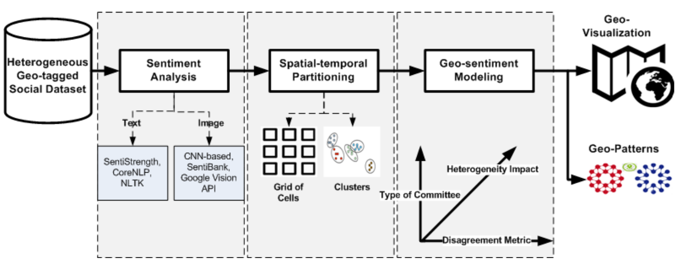

Fig. 1. Geo-Sentiment Framework

III. Geo-sentiment Analysis Framework

We propose a framework (see Fig. 1) to analyze the sentiment of geo-tagged social media data objects in disaster situations. Our framework consists of three phases; 1) sentiment analysis, 2) spatial-temporal partitioning, 3) geo-sentiment modeling. In what follows, we describe each phase in detail.

A. Sentiment Analysis Phase

In this phase, we employ a set of well-known sentiment classifiers on each object to extract the sentiment contained in the object. In our framework, the utilized text sentiment classifiers include SentiStrength, CoreNLP, and NLTK while the visual sentiment classifiers include the CNN-based classifier proposed by Campos et al., SentiBank and Google Vision API. SentiBank extracts 1200 adjective-noun pairs (ANP) for each image. To generate sentiment labels, we run a Naïve classifier where each of the detected ANPs has a sentiment score associated with it, and for each image, we compute the average of the top 10 detected ANPs. Despite the sentiment prediction limitation by Google Vision API, it is used to analyze the content of an image and extract a set of descriptive keywords to be fed to the text sentiment classifiers; thus the sentiment label of the image becomes the average of the sentiment score generated by the text classifiers.

As shown in Table I, each sentiment classifier has its scale to represent its results, which makes it hard to understand and straightforwardly compare the results across classifiers. Hence, our framework maps the generated sentiment scores from different classifiers into three uniform labels: negative (-1), neutral (0), and positive (1).

TABLE I. Sentiment Score Mapping Among Various Sentiment Classifiers

Classifier	Ranges		
	Negative (-1)	***Neutral (0)***	***Positive (1)***
SentiStrength	{-4,-3,-2,-1}	0	{1,2,3,4}
CoreNLP	{0, 1}	2	{3, 4}
NLTK	[-1, -0.1]	(-0.1, 0.1)	[0.1, 1]
CNN	[-1, -0.1]	(-0.1, 0.1)	[0.1, 1]
SentiBank	[-4, -1]	(-1, 1)	[1, 4]
Google Vision API	[-1, -0.1]	(-0.1, 0.1)	[0.1, 1]

B. Spatial-temporal Partitioning Phase

The impact of a disaster depends on time and location. Hence, to analyze a disaster situation thoroughly and reliably, we partition the data objects obtained from social media first temporally and then spatially. For temporal partitioning, data is split into a set of time windows (e.g., days in hurricane case or hours in earthquake case), the length of each time window can be configured to suit the type and speed of a disaster. Next, the data of each time window is partitioned spatially. There are different mechanisms to partition data (e.g., space-based or clustering-based). In the space-based mechanism, the global geographical area is partitioned into disjointed sub-areas by using a space partitioning technique such as uniform grid, quad-tree, kd-Tree [23], or Voronoi diagram [24]. Meanwhile, clustering-based mechanism creates partitions (i.e., clusters) of data objects based on the relative proximity of the point locations to each other. Examples of clustering techniques include K-Means [3] and DBSCAN [4].

Note that in our experiments, the framework partitions data objects into day-length windows and then the data in each window are partitioned by using either Grid or K-means. This phase generates a set of local geographical regions (i.e., corresponding to cells or clusters) for further analysis.

C. Geo-sentiment Modeling Phase

After generating a set of disjointed local geographical regions by the spatial-temporal partitioning phase, we propose a geo-sentiment model for estimating a sentiment score for a local region. This model is to address the challenges which we discussed earlier by considering three factors when estimating the region's overall sentiment: a) *the type of sentimental committee per object* which evaluates the certainty of the object's sentiment, b) *disagreement metric* which measures the sentiment inconsistency among the sentimental committee, and, c) *Data heterogeneity* which considers the source of the object and the importance of the object type in sentiment analysis.

The type of sentimental committee per object has three variants: $\mathbb{C}_C$, $\mathbb{C}_N$, or hybrid committee comprising both $\mathbb{C}_C$ and $\mathbb{C}_N$ (referenced as $\mathbb{C}_H$). With $\mathbb{C}_C$, the sentiment certainty of an object is determined by a set of sentiment classifiers; thus, the purpose of the classifiers committee is to select the objects whose sentiments are estimated with a high certainty. Meanwhile, with $\mathbb{C}_N$, the sentiment certainty of an individual object is assessed by that of the objects in its vicinity. In particular, when retrieving $\mathbb{C}_N$ of an O_i, all objects within a given spatial proximity (referred to as σ) are retrieved even if they exist outside the object's region. Consequently, $\mathbb{C}_N$ not only supports estimating the object's sentiment with high certainty among their neighbors but also smooths the predicted sentiments of the neighboring local regions. With $\mathbb{C}_H$, the committee of an object embodies the sentiment labels generated by a set of classifiers for the object itself and its geo-neighbors. Thus, $\mathbb{C}_H$ can overcome both the classification and geo-sentiment discrepancy challenges.

Each type of sentimental committee per object includes a set of sentiment labels which are used to generate a sentiment score for the target object with a certain certainty threshold (referred to as ε). We quantify the certainty by evaluating the level of disagreement among a committee of labels for each object. For this sake, our framework utilizes two distinct metrics: variance and entropy. Variance (defined in Eq. 1) is a statistical measure which captures the spread quantitatively among a set of values around their mean. Meanwhile, entropy (defined in Eq. 2) is a probabilistic measure which captures the logarithm of the probability distribution.

$$\sigma^2 = \sum_{i=1}^{n}(x_i - \mu)^2 / n \tag{1}$$

$$Entropy(o) = -\sum_t \frac{V(t,G)}{|G|} * \log\frac{V(t,G)}{|G|} \tag{2}$$

In the variance equation, μ denotes to the mean of sentiment scores. In the entropy equation, t denotes to the class of a sentiment label (i.e., positive, negative, or neutral), G refers to the object's committee (i.e., $\mathbb{C}_C$, $\mathbb{C}_N$, or $\mathbb{C}_H$), and $V(t, G)$ is the number of elements in G which are labeled by t. The entropy is maximized when the committee's elements are distributed evenly among the sentiment classes, and it is minimized when all elements are classified into one sentiment class. Since entropy values increase along with the size of the committee, we normalize the entropy value of each object by the maximum possible entropy (i.e., log |G|).

The disaster-related dataset obtained from different social media can be either single-type object (text or image objects) or heterogeneous multimedia objects (i.e., both text and image objects). When analyzing a heterogeneous dataset, the aggregated sentiment of a local geographical region is a weighted average of the sentiment scores of the region's objects where the object's sentiment is weighted by a constant based on the object type (α is a constant parameter for weighting the sentiment of image objects, and β is another constant parameter for text objects; $0 \leq \alpha, \beta \leq 1$, $\alpha + \beta = 1$). The value of weight parameters can be interpreted in different ways; the importance of the object type (i.e., how the object content reflects the disaster situation), the importance of the social media which provides the object, or the ratio of each object type among all objects. In this paper, we consider that the weight parameters represent the importance of the object type. The values of these parameters are set heuristically.

The procedure of our model is illustrated in Algorithm 1.

Algorithm 1: Geo-Sentiment Model

Input: Set (R) of local geographical regions
H = isHeterogeneous(R.O)
for each $r \in R$
  for each $o \in r.O$
    G = Committee (o)
    Disagreement Metric: DM(o) = Variance(G) or Entropy (G)
    $Sentiment(o) = \overline{Sentiment(G)}$
  end for
  Sentiment Certainty (r): $C = 1 - \overline{DM(r.O)}$
  if (C > ε)
    if (H = *true*)
      $Sentimen(r) = \overline{weight * sentiment(r.O)}$
    else
      $Sentimen(r) = \overline{Sentiment(r.O)}$
  else
    Discard the sentiment of r
  end if
end for

IV. Experimental Results

A. Experimental Methodology

Case-study Datasets: We considered two specific disaster cases: Hurricane Sandy (referred to as *SANDY*) and Napa Earthquake (referred to as *NAPA*). For *SANDY*, we collected tweets and images posted on Twitter and Flickr, respectively, during the period from October 22, 2012 (the day Sandy was formed) till November 2, 2012 (the day Sandy dissipated) in the area of the Northeastern United States. Our hypothesis is that people's negative sentiment is correlated with the progress of a disaster. Thus, we selected this specific 12-day period to understand how people's emotion changed while a hurricane passed by using sentiment analysis of social media data in the disaster affected area. The most affected areas by Sandy are New York and New Jersey and the storm hit New York City on

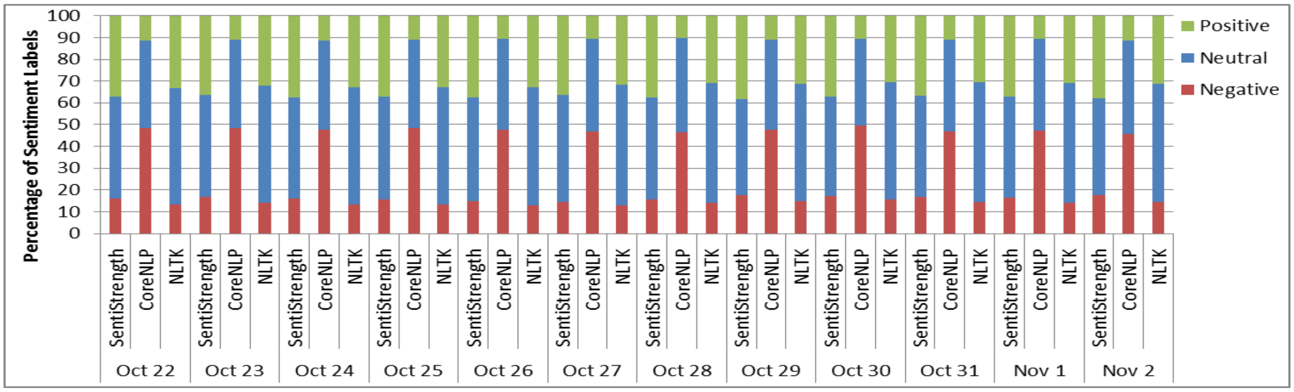

Fig. 2. Percentage of Sentiment Labels Based on Different Text Sentiment Classifiers in the period Oct 22-Nov2, 2012 (*SANDY-Tweets*)

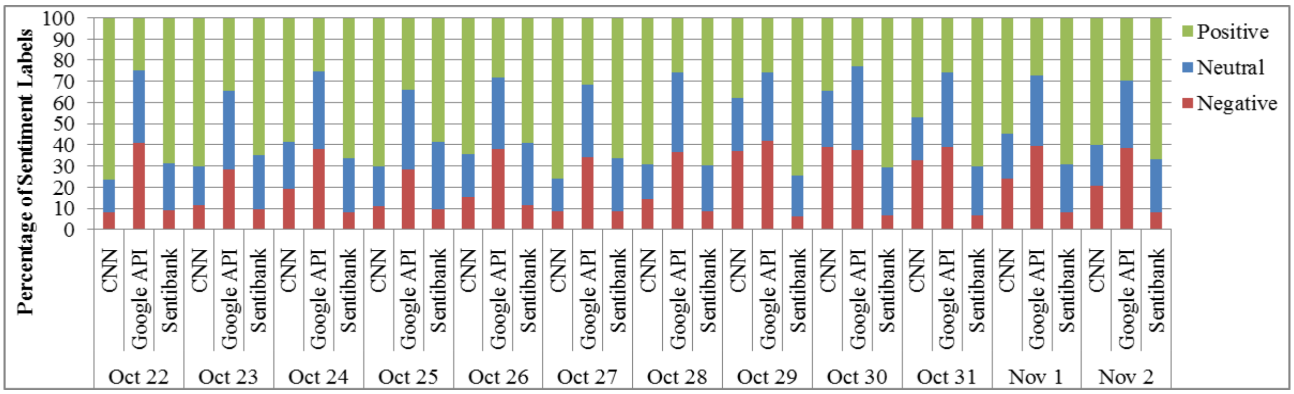

Fig. 3. Percentage of Sentiment Labels Based on Different Visual Sentiment Classifiers in the period Oct 22-Nov2, 2012 (*SANDY-Images*)

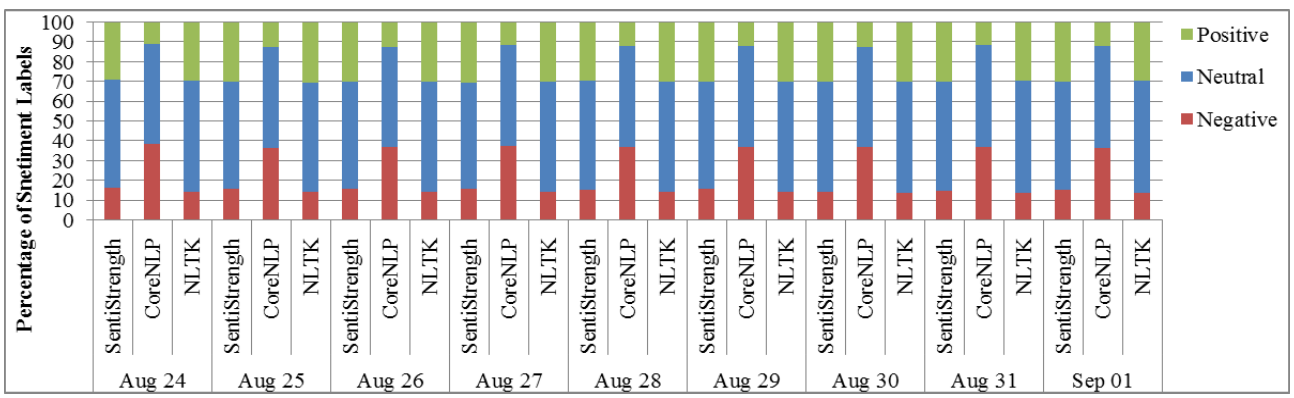

Fig. 4. Percentage of Sentiment Labels Based on Different Visual Sentiment Classifiers in the period Aug 24-Sep 1, 2014 (*NAPA-Tweets*)

October 29, 2012 [44]. To collect Sandy Twitter dataset, we used the public dataset released by Wang et al. [39] which contains a list of IDs of around 6 million geo-tagged tweets. Using these Tweet IDs, we crawled all necessary information of the Twitter posts (i.e., timestamp, geo-location, tweet message, and user ID) with Twitter Streaming API [36]. This list of Twitter posts is referred to as *SANDY-Tweets*. Based on the spatial-temporal coverage of the *SANDY-Tweets* dataset, we crawled all public geo-tagged images through both Flickr Photo Search [37] and GetInfo [38] APIs; the resulting list of Flickr images is referred to as *SANDY-Images*. Regarding *NAPA*, we used the Twitter dataset released by Pfeffer et al. [40] which contains a list of IDs of geo-tagged tweets posted within the United States during the period of June 1, 2014 till November 30, 2014. Then, we extracted only tweet IDs in the period of August 24, 2014 (the day Napa earthquake happened) until September 1, 2014 which were taken in the North San Francisco Bay Area. Subsequently, we crawled Twitter posts using Twitter streaming API to construct the *NAPA-Tweets* dataset.

For *SANDY-Tweets* and *NAPA-Tweets*, we removed the tweets which were generated by spamming or non-human bots since such tweets may falsify the results of our analysis. Following the technique by To et al. [21], we eliminated tweets of a user who generates more than 15 tweets per day. A summary of our datasets is described in Table II.

Each object of *SANDY-Tweets* and *NAPA-Tweets* is associated with three sentiment labels generated by SentiStrength, CoreNLP, and NLTK. Meanwhile, each object of *SANDY-Images* is associated with three sentiment labels generated by CNN, SentiBank, and Google Vision API[1]. Figs. 2, 3, and 4 show the percentage of sentiment labels per day based on multiple sentiment classifiers with *SANDY-Tweets*, *SANDY-Images*, and *NAPA-Tweets*, respectively. The percentage difference of sentiment labels by different sentiment classifiers demonstrates the classification discrepancy. Without the geo-sentiment analysis, we notice that the majority of tweet objects were labeled as neutral while the majority of image objects were labeled as positive. The general positive trend in images is explained by the fact that people usually post images on social media when they are in a positive mood. Meanwhile, text posts on social media usually represent a general talk; thus the majority of tweets are neutral.

Used Techniques and Parameter Values: We evaluated our framework by varying the techniques used in the spatial-temporal partitioning (i.e., Phase 2) and geo-sentiment modeling phase (i.e., Phase 3) as shown in Table III. We also varied the framework parameters (see Table IV). The default technique and parameter values are underlined.

TABLE II. DATASET STATISTICS

		SANDY	*NAPA*
Tweets	***Total # of tweets***	2.7M	406K
	Avg. # of Tweets per Day	223K	45K
	Avg. [2] # of Tweets per 10*10 km²	351	455
Images	***Total # of Images***	84K	-
	Avg. # of Images per Day	7K	-
	Avg. [2] # of Images per 10*10 km²	27	-

TABLE III. FRAMEWORK TECHNIQUES

Partition Technique	<u>Grid</u>, Clustering
Sentimental Committee Type	$\mathbb{C}_N$, $\mathbb{C}_C$, $\underline{\mathbb{C}_H}$
Disagreement Measure	Variance, <u>Entropy</u>
Types of Objects	Tweets, Images, <u>Heterogeneous</u>

TABLE IV. FRAMEWORK PARAMETERS

ε	0.3, <u>0.5</u>, 0.7, 0.9
α	0.4, 0.5, <u>0.6</u>, 0.7
σ	0.125, 0.25, 0.5, <u>1.0</u>, 2.0, 4.0 km

Evaluation of Framework: To evaluate our framework, we implemented a simplified version (referred to as *BASELINE*) of our framework where it uses one sentiment classifier at Phase 1, the Grid technique for spatial partitioning at Phase 2, averaging the sentiment labels associated with the objects in each local region (i.e., there is no sentimental committee type or disagreement measure). We assumed that *BASELINE* is one of the state-of-the-art geo-spatial sentiment analysis methods and compared our new approaches with it. Furthermore, we validate the analysis results of our framework with the reports provided by FEMA and USGS (only in the case of Napa earthquake).

[1] We plan to publish our dataset containing the sentiment scores when the paper is accepted.

[2] We discarded the empty areas when calculating the average.

B. *Discrepancy among Sentiment Classifiers*

In this set of experiments, we show the various sentiment analysis results of *BASELINE* with *SANDY-Tweets and SANDY-Images*. Figs. 5, 6, and 7 show the percentage of three resulting sentiment labels from all local regions (i.e., grid cells) using *BASELINE* with the text sentiment classifiers SentiStrength, CoreNLP, and NLTK, respectively. The majority of local regions using both SentiStrength and NLTK were positive while the results were inverted using CoreNLP. Similarly, as shown in Figs. 8, 9, and 10, *BASELINE* with different visual sentiment classifiers produced very different results. These clearly demonstrate the discrepancy of sentiment classifications which yields to contradicting analysis results with the same dataset. Furthermore, the results of sentiment analysis of *BASELINE* do not show the trend of negative sentiment labels during the disaster. Consequently, this finding supports the importance of the sentiment analysis phase in our framework which employs multiple classifiers to overcome the biased results of an individual classifier.

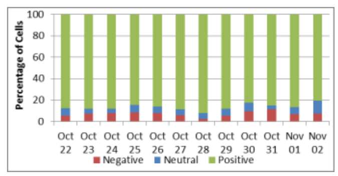

Fig. 5. Percentage of Cells Among Sentiment Labels using *BASELINE* (SentiStrength) w/ *SANDY-Tweets*

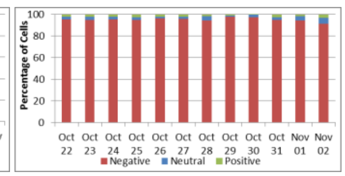

Fig. 6. Percentage of Cells Among Sentiment Labels using *BASELINE* (CoreNLP) w/ *SANDY-Tweets*

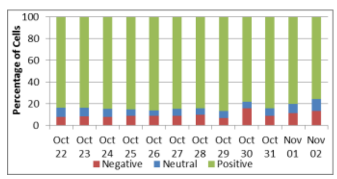

Fig. 7. Percentage of Cells Among Sentiment Labels using *BASELINE* NLTK) w/ *SANDY-Tweets*

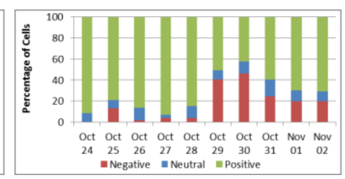

Fig. 8. Percentage of Cells Among Sentiment Labels using *BASELINE* CNN) w/ *SANDY-Images*

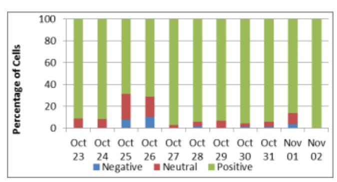

Fig. 9. Percentage of Cells Among Sentiment Labels using *BASELINE* (SentiBank) w/ *SANDY-Images*

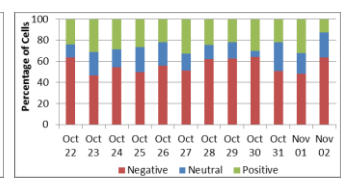

Fig. 10. Percentage of Cells Among Sentiment Labels using *BASELINE* (Google Vision API) w/ *SANDY-Images*

C. *The impact of Geo-sentiment Modeling Phase*

1) *Sentimental Committee Type*

The impact of $\mathbb{C}_N$: In this set of experiments, we used a single individual sentiment classifier at Phase 1 in our framework and $\mathbb{C}_N$ at Phase 3 to understand the impact of $\mathbb{C}_N$. Figs. 11 to 16 show the distribution of local regions among sentiment labels using SentiStrength, CoreNLP, NLTK, CNN, SentiBank, and Google Vision API, respectively. When comparing Figs. 11-16 with their corresponding Figs. 5-10, the percentage of local regions which are labeled as neutral is reduced because of employing $\mathbb{C}_N$. Given that local regions labeled neutral sentiment indicates that either all objects in that region has a neutral sentiment or there is a randomness of sentiment labels of the region objects. The latter case depicts geo-neighbors discrepancy; thus the use of $\mathbb{C}_N$ reduces its impact.

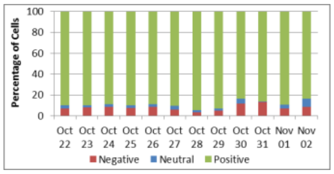

Fig. 11. Percentage of Cells Among Sentiment Labels using $\mathbb{C}_N$ - SentiStrength w/ *SANDY-Tweets*

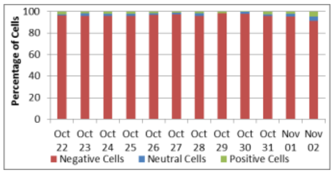

Fig. 12. Percentage of Cells Among Sentiment Labels using $\mathbb{C}_N$ - CoreNLP w/ *SANDY-Tweets*

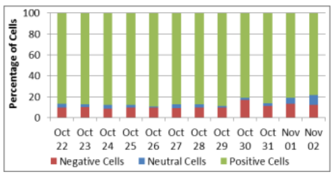

Fig. 13. Percentage of Cells Among Sentiment Labels using $\mathbb{C}_N$-NLTK w/ *SANDY-Tweets*

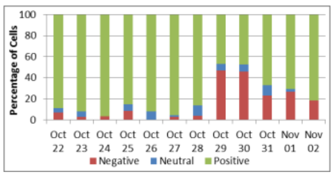

Fig. 14. Percentage of Cells Among Sentiment Labels using $\mathbb{C}_N$ - CNN w/ *SANDY-Images*

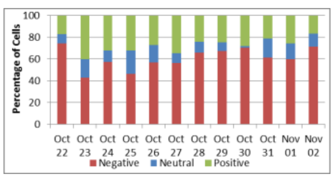

Fig. 15. Percentage of Cells Among Sentiment Labels using ℂN - SentiBank w/ SANDY-Images

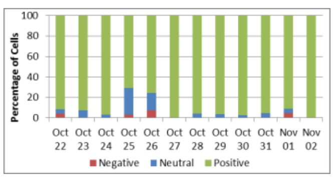

Fig. 16. Percentage of Cells Among Sentiment Labels using ℂN – Google Vision API w/ SANDY-Images

The impact of $\mathbb{C}_C$***:*** In this set of experiments, we used multiple sentiment classifiers at Phase 1 in our framework and $\mathbb{C}_C$ at Phase 3 to understand the impact of $\mathbb{C}_N$. Figs. 17 and 18 show the distribution of local regions among sentiment labels with *SANDY-Tweets* and *SANDY-Images*, respectively. The method of merging the results of multiple sentiment classifiers and selecting objects whose sentiment labels are certain minimizes the impact of classification discrepancy; hence the sentiment analysis becomes consistent. In Fig. 17, the percentage of negative regions is signified in the analysis of *Sandy-Tweets.* Meanwhile, with *Sandy-Images*, the proportion of negative local regions becomes evident, especially after the disaster happened (i.e., after Oct 29, 2012). With $\mathbb{C}_C$, there are still neutral local regions which can be avoided by using $\mathbb{C}_N$.

The impact of $\mathbb{C}_H$***:*** In this set of experiments, we used multiple sentiment classifiers at Phase 1 in our framework and $\mathbb{C}_H$ at Phase 3 to understand the impact of $\mathbb{C}_H$. Figs. 19 and 20 show the distribution of local regions among sentiment labels with *SANDY-Tweets* and *SANDY-Images*, respectively. Combing $\mathbb{C}_N$ with $\mathbb{C}_C$ enables selecting the objects whose sentiment labels are certain based on multiple sentiment classifiers and the sentiment of their geo-neighbors. As shown in Fig. 19, the percentage of negative regions reaches to its highest value (i.e., 63%) on Oct. 30, 2012 using *SANDY-Tweets*, while the proportion of negative regions (see Fig. 20) reaches to its highest value (i.e., 15%) on Oct. 29, 2012 using *SANDY-Images*. With $\mathbb{C}_H$, the framework combines the advantages gained by both $\mathbb{C}_C$ and $\mathbb{C}_N$; hence avoids both classification and geo-neighbors discrepancies.

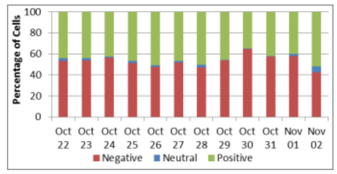

Fig. 17. Percentage of Cells Among Sentiment Labels using $\mathbb{C}_C$ w/ *SANDY-Tweets*

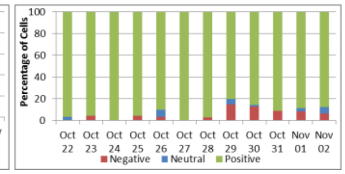

Fig. 18. Percentage of Cells Among Sentiment Labels using $\mathbb{C}_C$ w/ *SANDY-Images*

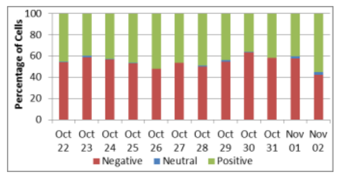

Fig. 19. Percentage of Cells Among Sentiment Labels using $\mathbb{C}_H$ w/ *SANDY-Tweets*

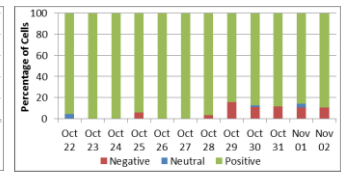

Fig. 20. Percentage of Cells Among Sentiment Labels using $\mathbb{C}_H$ w/ *SANDY-Images*

2) Varying Disagreement Measure

In this set of experiment, we show the impact of the chosen disagreement measure in our framework. Figs. 21 and 22 show the distribution of local regions among sentiment labels using Variance measure with *SANDY-Tweets* and *SANDY-Images*, respectively. When comparing Figs. 21 and 22 (Variance) with their corresponding Figs. 19 and 20 (Entropy), we have two observations: a) the percentage of neutral local regions increased using Variance than that with Entropy, which implies that the impact of $\mathbb{C}_N$ decreased with Variance, and b) however the Variance measure was able to show the negative trend with *SANDY-Tweet*s while Variance fails to demonstrate the negative trend of Sandy-Images especially when the most damage by Hurricane Sandy happened in New York area (i.e., after October 29, 2012). This concludes that Entropy is a more stable disagreement measure than Variance.

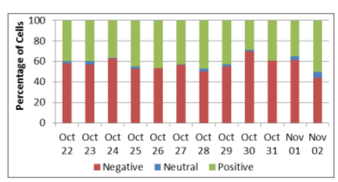

Fig. 21. Percentage of Cells Among Sentiment Labels using Variance w/ *SANDY-Tweets*

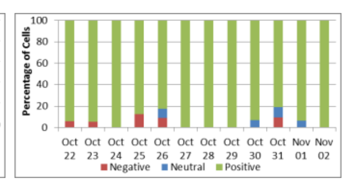

Fig. 22. Percentage of Cells Among Sentiment Labels using Variance w/ *SANDY-Images*

3) Varying Object Types

In this set of experiments, we show the impact of using heterogeneous object types to detect disaster situation in a geographical area through our framework. Fig. 23 shows the distribution of local regions among sentiment labels with heterogeneous datasets of both *SANDY-Tweets* and *SANDY-Images*. When comparing Fig. 23 (*SANDY-Heterogeneous*) with Figs. 19 (*SANDY-Tweets*) and Fig. 20 (*SANDY-Images*) we observe that geo-sentiment analysis with the heterogeneous dataset was superior due to two reasons: a) obtaining more objects in each local region which increase the effect of $\mathbb{C}_H$, and b) the usage of combined weighted aggregated sentiment score. In Fig. 23, the percentage of negative local regions reached to the highest value (i.e., 80%) on Oct. 30, 2012.

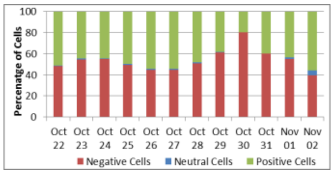

Fig. 23. Percentage of Cells Among Sentiment Labels using Grid w/ *SANDY-Heterogeneous*

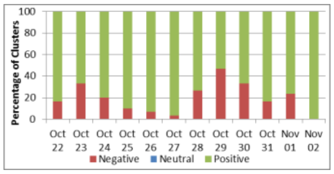

Fig. 24. Percentage of Clusters Among Sentiment Labels using Clustering w/ *SANDY-Heterogeneous*

D. *Varying Spatial-temporal Partitioning Technique*

In this set of experiments, we show the impact of partitioning techniques (Grid or Clustering) at Phase 2 in our framework. Fig. 24 shows the distribution of local regions among sentiment labels using clustering with *SANDY-Heterogeneous*. When comparing Fig. 24 (Clustering) with Figs. 23 (Grid), we observe that both partitioning techniques were able to detect the disaster situation, however, with the Grid partitioning technique, the percentage of negative local regions was signified compared to that with Clustering. With Grid, the analyzer predefines the spatial coverage of each cell (i.e., local region) while with clustering the spatial coverage of each cluster (i.e., local region) is dynamically determined; hence it is possible that the spatial coverage area of a cluster can be larger than that of a grid cell. Given that, the accuracy of sentiment analysis increases with small local regions; the Grid-based analysis was able to detect the geo-sentiment of local regions accurately.

E. *Geo-sentiment Analysis Framework vs. Ground Truth*

Here, we compare the analysis result of our framework with the reports generated by FEMA and USGS.

Hurricane Sandy Case: Fig. 25 shows the visualization of the results of our geo-sentiment analysis framework on Oct. 30, 2012 while Fig. 26 shows the map reported by FEMA [41] which declares the New York's counties (in Orange) that required public or governmental assistance during Hurricane Sandy. Our results were able to detect the negative trend of the New York counties which marked as areas which needed assistance based on FEMA report. Furthermore, our analysis shows that other coastal areas (e.g., New Jersey and Maryland), which were affected by Hurricane Sandy [44], had mostly negative sentiment labels.

Napa Earthquake Case: Fig. 27 shows the percentage of three resulting sentiment labels from all local regions using our framework with *NAPA*. The analysis result showed that the highest proportion (78%) of negative sentiment labels occurred on Aug. 24, 2014 (the day of Napa earthquake). The visualized result of our geo-sentiment analysis on Aug 24, 2014 is shown in Fig. 29 which coincides with both the FEMA' report (Fig. 30) [42] and the USGS' shakemap (Fig. 28) [43] of the Napa earthquake. Based on our analysis, the areas surrounding Napa Valley (e.g., San Francisco and Marin) also showed negative sentiment which potentially depicts people panic in their posted tweets however they were not part of the disaster situation (i.e., tweets does not necessarily reflect an event happening in people location).

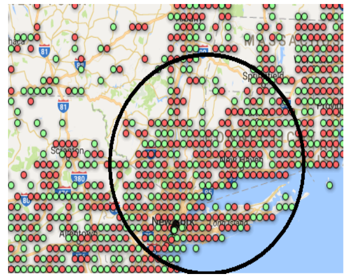

Fig. 25. Visualization of Geo-sentiment Analysis on Oct. 30, 2012 w/ *SANDY-Heterogeneous*

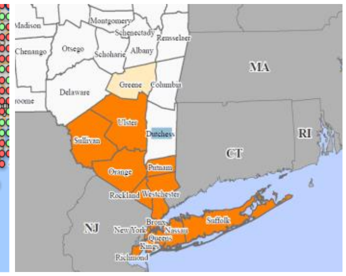

Fig. 26. New York's Counties which needed Assistance based on FEMA Report at the time of Hurricane Sandy

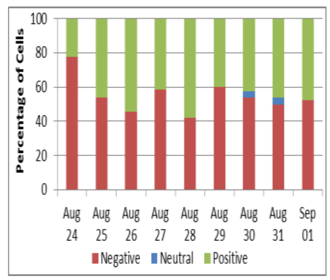

Fig. 27. Percentage of Cells among Sentiment Labels w/ *NAPA*

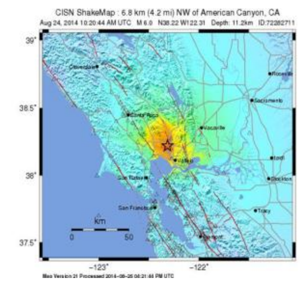

Fig. 28. Shakemap of Napa Earthquake based on USGS

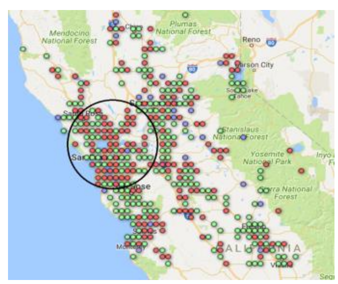

Fig. 29. Visualization of Geo-sentiment Analysis on Aug. 24, 2014 w/ *NAPA*

Fig. 30. California's Counties (in Pink) Which Needed Assistance based on FEMA Report at the Time of Napa Earthquake

F. *Varying Framework Parameters*

The impact of certainty threshold ε: Fig. 31 shows the average of resulting negative labels for local regions when varying ε. We chose to show the percentage of negative sentiment labels because the negative sentiment class is the expected sentiment in disasters. Enlarging the value of ε ensures that the selected local regions convey the minimum degree of discrepancy of both geo-neighbors and classification; hence increases the accuracy of sentiment analysis obtained by the framework.

The impact of image weight parameter α: Fig. 32 shows the average of resulting negative labels for local regions when varying α. Decreasing the weight of image objects resulted in an increase in the percentage of negative sentiment labels. This indicates that people express the negative sentiment with words more than images.

The impact of spatialge range σ: Fig. 33 shows the average of resulting neutral labels for local regions when varying σ. We chose to show the percentage of neutral sentiment labels for local regions because they are potentially generated due to the sentiment randomness of individuals within a local region. Based on our experiments, increasing σ resulted with a decrease of neutral areas. The increase of σ leads to the increase of the potential geo-neighbors for estimating geo-neighbors discrepancy which reduces the sentiment randomness among neighbors; hence potentially decreases the neutral sentiment labels.

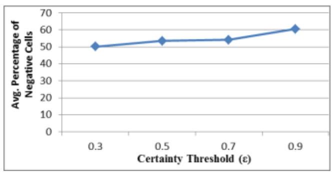

Fig. 31. Impact of ε on geo-sentiment analysis w/ *SANDY-Heterogeneous*

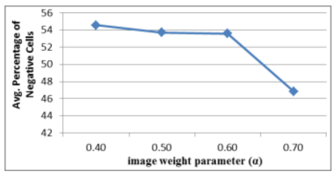

Fig. 32. Impact of α on geo-sentiment analysis w/ *SANDY-Heterogeneous*

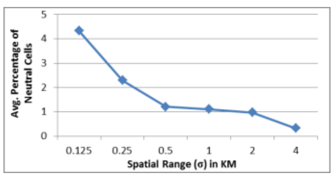

Fig. 33. Impact of σ on geo-sentiment analysis w/ *SANDY-Heterogeneous*

V. Related Work

Analyzing the public sentiment of people through social media during a disaster has been extensively studied. However, these studies focused on analyzing one source of social media [33] [45] [46] [47] [48]. Examples of research works related to Twitter-based sentiment analysis include: a) Nagy and Stamberger [45] investigated the sentiment of tweets using SentiWordNet classifier [30] during the fire caused by the San Bruno pipeline explosion on September 2010, b) Schulz et al. [47] proposed a tuned sentiment analysis model for disaster situation and evaluated their model on tweets during the Hurricane Sandy, c) Mandel et al. [48] devised sentiment classifier with a trained disaster-based model and assessed their classifier on Twitter messages related to Hurricane Irene (August 2011), and d) Borth et al. [24] proposed a visual sentiment classifier, SentiBank, and evaluated it on a set of image tweets collected during Hurricane Sandy. Besides, there is another group of studies which focused on sentiment analysis on Flickr (e.g., [46]) or Facebook (e.g., [33]) data. Our framework integrates a heterogeneous set of objects collected from different social media platforms for a robust geo-sentiment analysis. To the best of our knowledge, there is one study (Flaes et al. [32]) which has utilized sentiment analysis on different social media to predict city livability (i.e., not related to disaster situation). The work performed by Flaes et al. does not consider the geo-neighbor and classification discrepancy challenges which we addressed in our framework. Furthermore, the sentiment analysis proposed by Flaes et includes neither spatial partitioning nor multiple classifiers.

The machine learning community has investigated the classification discrepancy which results when utilizing multiple classifiers. One of the proposed techniques to measure the disagreement level among multiple classifiers is entropy-based [31]. Using the entropy measure, Lu et al. [29] employed multiple sentiment classifiers to analyze tweets related to Ebola disaster. The sentiment analysis performed by Lu et al. neither considers the geo properties of objects nor analyzes heterogeneous objects (i.e., both challenges of geo-neighbor discrepancy and dataset heterogeneity are not addressed).

VI. CONCLUSION

In this paper, we studied the geo-sentiment analysis of geo-tagged data objects obtained from heterogeneous social media in disasters. The geo-sentiment analysis comprises three challenges: discrepancy among multiple sentiment classifiers, discrepancy among the sentiments of geo-neighbors, and dataset heterogeneity. To overcome these challenges, we proposed a novel framework composed of three phases: sentiment analysis, spatial-temporal partitioning, and geo-sentiment modeling. To estimate the aggregated sentiment score for a set of objects in a local region, our geo-sentiment model considers the sentiment labels generated by multiple classifiers in addition to those of geo-neighbors. To obtain sentiment with high certainty, the model measures the disagreement among correlated sentiment labels either by entropy or variance metric. We used our framework to analyze the disasters of Hurricane Sandy and Napa Earthquake based on datasets collected from Twitter and Flickr. Our analysis results were analogous to FEMA, and USGS reports.

Our future work includes a) evaluating our geo-sentiment framework in applications other than disaster situation, b) extending our framework by utilizing a multi-modal sentiment classifier, and c) investigating other social media rather than Twitter and Flickr.

References

[1] Google Cloud Natural API, https://cloud.google.com/natural-language/.

[2] J.-C. Na, H. Sui, C. Khoo, S. Chan, and Y. Zhou, “Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews,” in Proc. of Conference of the International Society for Knowledge Organization (ISKO), pp. 49–54, 2004.

[3] J. MacQueen, "Some methods for classification and analysis of multivariate observations," In Proc. of the fifth Berkeley symposium on mathematical statistics and probability, vol. 1, no. 14, pp. 281-297. 1967.

[4] M. Ester, H. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," In Proc. of Int. conference on Knowledge Discovery and Data Mining (KDD), Portland, Oregon, USA, pp. 226-231. 1996.

[5] Microsoft Azure Computer Vision API, https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/.

[6] Google Cloud Vision API, https://cloud.google.com/vision/.

[7] J. Yuan, S. Mcdonough, Q. You, and J. Luo. "Sentribute: image sentiment analysis from a mid-level perspective." In Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining, p. 10. ACM, 2013.

[8] L. Mitchell, M. R. Frank, K. D. Harris, P. S. Dodds, and C. M. Danforth. "The geography of happiness: Connecting twitter sentiment and

expression, demographics, and objective characteristics of place." PloS one 8, no. 5 (2013): e64417.

[9] J. Jia, S. Wu, X. Wang, P. Hu, L. Cai, and J. Tang. "Can we understand van gogh's mood?: learning to infer affects from images in social networks." In Proceedings of the 20th ACM international conference on Multimedia, pp. 857-860. ACM, 2012.

[10] M. Thelwall, K. Buckley, and G. Paltoglou.. "Sentiment strength detection for the social web." Journal of the American Society for Information Science and Technology 63, no. 1 (2012): 163-173.

[11] V. Campos, B. Jou, and X. G.. "From Pixels to Sentiment: Fine-tuning CNNs for Visual Sentiment Prediction." Image and Vision Computing (2017).

[12] The IBM Watson AlchemyLanguage service, https://www.ibm.com/watson/developercloud/alchemy-language.html.

[13] S. Bird. "NLTK: the natural language toolkit." In Proceedings of the COLING/ACL on Interactive presentation sessions, pp. 69-72. Association for Computational Linguistics, 2006.

[14] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury. "Twitter power: Tweets as electronic word of mouth." Journal of the American society for information science and technology 60, no. 11 (2009): 2169-2188.

[15] A. Kongthon, C. Haruechaiyasak, J. Pailai, and S. Kongyoung. "The role of Twitter during a natural disaster: Case study of 2011 Thai Flood." In 2012 Proceedings of PICMET'12: Technology Management for Emerging Technologies, pp. 2227-2232. IEEE, 2012.

[16] S. Wakamiya, R. Lee, and K. Sumiya. "Crowd-based urban characterization: extracting crowd behavioral patterns in urban areas from twitter." In Proceedings of the 3rd ACM SIGSPATIAL international workshop on location-based social networks, pp. 77-84. ACM, 2011.

[17] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment." ICWSM 10 (2010): 178-185.

[18] M. J. Paul and M. Dredze. "You are what you Tweet: Analyzing Twitter for public health." ICWSM 20 (2011): 265-272.

[19] Y. Liu, X. Huang, A. An, and X. Yu. "ARSA: a sentiment-aware model for predicting sales performance using blogs." In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 607-614. ACM, 2007.

[20] T. Sakaki, M. Okazaki, and Y. Matsuo. "Earthquake shakes Twitter users: real-time event detection by social sensors." In Proceedings of the 19th international conference on World wide web, pp. 851-860. ACM, 2010.

[21] H. To, S. Agrawal, S. H. Kim, C. Shahabi. "On Identifying Disaster-Related Tweets: Matching-based or Learning-based?" In Proc. Int. Conf. on Multimedia Big Data, 2017.

[22] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan. "A system for real-time twitter sentiment analysis of 2012 us presidential election cycle." In Proceedings of the ACL 2012 System Demonstrations, pp. 115-120. Association for Computational Linguistics, 2012.

[23] S. Siersdorfer, E. Minack, F. Deng, and J. Hare. "Analyzing and predicting sentiment of images on the social web." In Proceedings of the 18th ACM international conference on Multimedia, pp. 715-718. ACM, 2010.

[24] D. Borth, R. Ji, T. Chen, T. Breuel, and Shih-Fu Chang. "Large-scale visual sentiment ontology and detectors using adjective noun pairs." In Proceedings of the 21st ACM international conference on Multimedia, pp. 223-232. ACM, 2013.

[25] J. Krauss, S. Nann, D. Simon, P. A. Gloor, and K. Fischbach. "Predicting Movie Success and Academy Awards through Sentiment and Social Network Analysis." In ECIS, pp. 2026-2037. 2008.

[26] C. Xu, S. Cetintas, K. Lee, and L. Li. "Visual sentiment prediction with deep convolutional neural networks." arXiv preprint arXiv:1411.5731 (2014).

[27] Q. You, J. Luo, H. Jin, and J. Yang. "Robust image sentiment analysis using progressively trained and domain transferred deep networks." arXiv preprint arXiv:1509.06041 (2015).

[28] Microsoft Azure Text Analytics API, https://azure.microsoft.com/en-us/services/cognitive-services/text-analytics/.

[29] Y. Lu, X. Hu, F. Wang, S. Kumar, H. Liu, and R. Maciejewski. "Visualizing social media sentiment in disaster scenarios." In Proc. of the 24th International Conference on World Wide Web, pp. 1211-1215. ACM, 2015.

[30] A. Esuli and F. Sebastiani. "Sentiwordnet: A publicly available lexical resource for opinion mining." In Proc. of the International Conference on Language Resources and Evaluation (LREC), vol. 6, pp. 417-422. 2006.

[31] I. Dagan, and Sean P. Engelson. "Committee-based sampling for training probabilistic classifiers." In Proc. of the International Conference on Machine Learning (ICML), pp. 150-157. The Morgan Kaufmann series in machine learning,(San Francisco, CA, USA), 1995.

[32] J. B. Flaes, S. Rudinac, and M. Worring. "What Multimedia Sentiment Analysis Says About City Liveability." In European Conference on Information Retrieval, pp. 824-829. Springer International Publishing, 2016.

[33] P. Dewan, V. Bharadhwaj, A. Mithal, A. Suri, and P. Kumaraguru. "Visual Themes and Sentiment on Social Networks To Aid First Responders During Crisis Events." arXiv preprint arXiv:1610.07772 (2016).

[34] R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts. "Recursive deep models for semantic compositionality over a sentiment treebank." In Proceedings of the conference on empirical methods in natural language processing (EMNLP), vol. 1631, p. 1642. 2013.

[35] S. E. Vieweg, "Situational awareness in mass emergency: A behavioral and linguistic analysis of microblogged communications." 2012.

[36] Twitter Streaming API, https://dev.twitter.com/streaming/overview.

[37] Flickr Photo Search API, https://www.flickr.com/services/api/flickr.photos.search.html.

[38] Flickr Photo GetInfo API, https://www.flickr.com/services/api/flickr.photos.getInfo.html.

[39] H. Wang, E. Hovy, and M. Dredze. "The hurricane sandy twitter corpus." In AAAI Workshop on the World Wide Web and Public Health Intelligence, pp. 20-24. 2015.

[40] J. Pfeffer and F. Morstatter. "Geotagged twitter posts from the united states: A tweet collection to investigate representativeness." Version: 1. GESIS Data Archive. Dataset. http://doi.org/10.7802/1166, 2016.

[41] New York Hurricane Sandy (DR-4085), https://www.fema.gov/disaster/4085.

[42] California Earthquake (DR-4193), https://www.fema.gov/disaster/4193.

[43] M 6.0 - 6km NW of American Canyon, California, https://earthquake.usgs.gov/earthquakes/eventpage/nc72282711.

[44] E. S. Blake, T. B. Kimberlain, J. Robert, J. P. Cangialosi, J. L. Beven II, National Hurricane Center (February 12, 2013). Hurricane Sandy: October 22 – 29, 2012 (Tropical Cyclone Report). United States National Oceanic and Atmospheric Administration's National Weather Service. Archived from the original (PDF) on February 17, 2013. Retrieved February 17, 2013

[45] A. Nagy, and J. Stamberger. "Crowd sentiment detection during disasters and crises." In Proc. of International Conference on Information Systems for Crisis Response And Management (ISCRAM), pp. 1-9, Vancouver, Canada. 2012.

[46] Y. C. Chien, A. Comber, and S. Carver. "Does Flickr work in disaster management?–a case study of Typhoon Morakot in Taiwan." (2017).

[47] A. Schulz, T. Thanh, H. Paulheim, and I. Schweizer. "A fine-grained sentiment analysis approach for detecting crisis related microposts." In Proc. of International Conference on Information Systems for Crisis Response And Management (ISCRAM), 2013.

[48] B. Mandel, A. Culotta, J. Boulahanis, D. Stark, B. Lewis, and J. Rodrigue. "A demographic analysis of online sentiment during hurricane irene." In Proc. of the Second Workshop on Language in Social Media, pp. 27-36. Association for Computational Linguistics, 2012.