Disaster-Related Participant Tweet Identification Using SVM

Lemuel John Beduya
University of the Philippines Cebu
Gorordo Ave., Lahug, Cebu City, Cebu
233 8203
lemuel.beduya@gmail.com

Kurt Junshean Espinosa
University of the Philippines Cebu
Gorordo Ave., Lahug, Cebu City, Cebu
233 8203
kpespinosa@up.edu.ph

ABSTRACT
The Philippines is a hot bed of disasters: earthquakes, flooding, and fires often occur in the country Moreover, in this part of the globe, majority of the populace are very attuned to social media with almost everyone who are either Twitter or Facebook users. This study took advantage of that and used Twitter in identifying the disaster-related participant tweets by Filipino users in the Philippines. This study will aid the Philippine government and other concerned organizations in their disaster management plans. In view of this, a multi-level binary classification on tweets was implemented using SVM. Specifically, the first level identified a tweet if it is a disaster-related participant tweet or not. The next level identified the type of disaster the participant is experiencing which can be flood, earthquake or others (fire, landslide, etc). In order to yield the best model for each data set, a 10-fold cross-validation was performed. The process yielded a model for each data set with an F_1 score of 0.73, 0.83 and 0.72 for disaster-related, flood-related and earthquake-related participant tweets respectively. The results of the study showed that it is indeed possible to identify participant tweets of any type of disaster in Twitter using SVM. Furthermore, this study can be used as a starting point in examining if it is possible to identify the disaster-prone areas in the Philippines using Twitter.

Categories and Subject Descriptors
I.2.7: [**Artificial Intelligence**]: Natural Language Processing – *Text analysis*.

General Terms
Algorithms, Measurement, Performance, Experimentation, Human Factors.

Keywords
Disaster, Tweet Classification, SVM, Machine Learning, Twitter, Social Media, Bag-of-words Model, tf-idf.

1. INTRODUCTION
The Philippines is a hot bed of disasters. Earthquakes often occur in the country which are caused by the volcanoes and earthquake generators that surrounds the archipelago since it is situated along the western part of the Pacific Ring of Fire [21]. One of which was the recent 7.2 magnitude earthquake in the Central Visayas which crumbled down many old churches in Bohol. It was reported that it was the strongest earthquake Philippines had in 23 years and caused a total of ₱2.25 billion worth of damages in Bohol, the quake's epicenter, and Cebu [1].

Many Filipinos cannot also forget that three weeks after the devastating earthquake, another disaster struck the Philippines again – the Super Typhoon Haiyan (Yolanda). Cyclones visit the country frequently and every year, around 19 tropical cyclones enter the Philippine Area of Responsibility (PAR). Out of these, 6 to 9 make landfall [25]. However, Super Typhoon Haiyan (Yolanda) was different, it was the strongest typhoon the country ever had, hence, called "super typhoon". It is said to be the equivalent of a Category 5 storm and caused catastrophic destruction in the Visayas, especially in Samar and Leyte, the hardest hit regions by the super typhoon. This apocalyptic super typhoon is said to have caused ₱39.82 billion worth of damages to infrastructure and agriculture according to National Disaster Risk Reduction and Management Council (NDRRMC)'s assessment [24].

Aside from natural disasters, sudden fires in the cities also frequently happen in the country. Most of the times, these fires are caused by human carelessness [21]. The country is indeed a hotbed of disasters.

Moreover, the Philippines is a country that is very attuned to social media [12]. The country is even nicknamed as "the social networking capital of the world" [18]. It is also the country with the highest social networking penetration in the Asia-Pacific region [17]. Although Twitter is only the second most popular social networking in the country following Facebook, the Philippines ranked 8th in countries with most Twitter users [18]. In fact, in the Philippines, even the government agencies have Twitter accounts to make dissemination of advisories or warnings to Filipinos faster and easier, such as the Department of Science and Technology (DOST)'s Twitter account @dost_pagasa which tweets weather forecasts and updates.

Indeed, there is a strong relationship between social media and disaster. In fact, there is a study which uses the tweets of Japanese Twitter users to detect if an influenza epidemic is happening and predict what type of influenza will spread in any given season [2]. Moreover, there also exists a study which considers the Twitter users as social sensors and uses their tweets to detect if an earthquake is currently occurring in Japan [19]. Aside from that, there is also a local study which uses tweets of Filipino Twitter users in understanding their behavior during a disaster [12]. Furthermore, there are also studies on identifying real world events such as disasters through Twitter [12]. Lastly, studies on how Twitter can contribute various categories of information to situational awareness during hazards also exist [26].

From these studies, the proponents realized that there is indeed a strong relationship between Twitter and disasters and it should be further studied. These studies encouraged the proponents to understand this relationship in the context of the disasters in the Philippines and the tweets of Filipino users about these disasters. Specifically, in this study, the proponents will identify the disaster-related participant tweets from Philippines that consequently may be used in identifying the disaster-prone areas in the Philippines. The proponents hope that this study will aid the concerned organizations (i.e. NDRRMC) and the government in its disaster management plans. Through this study, the proponents hope that that the loss the country suffers on every disaster will be minimized.

2. REVIEW OF RELATED LITERATURE

2.1 Social Media

In the recent years, social media experts have found the usefulness of social media as a great source of information which can be applied to business advertising and promotion of a specific social, cultural or political concern among others. Specifically, data from social networking sites have been used recently in disaster preparedness and disaster relief distribution operations which are relevant to this study. In fact, according to Gao, Barbier, and Goolsby, “Social media has recently played a critical role in natural disasters as an information propagator that can be leveraged for disaster relief.” [9]. A very good example of which is the Haiti earthquake in 2010 in which people shared and posted their personal experiences of the earthquake via Twitter, Flickr, Facebook, and Youtube. Their posts greatly helped Red Cross in properly distributing reliefs to the affected areas and in collecting donations for the victims. As a matter of fact, Red Cross received a total of US $8 million monetary donations for the victims of Haiti earthquake [21].

In Philippines, Facebook and Twitter are two of the Top 10 social networking sites in the country in 2010, being #1 and #4 respectively [23]. In fact, in the case of Facebook, Philippines is #8 in the countries with the most number of Facebook users (SocialBakers as cited in [13]). Thus, it is apparent that a large amount of data can be retrieved from Facebook and Twitter. Furthermore, studies show that these data can be definitely used in many ways such as disaster detection and sentiment analysis. One of which is Tolentino and Hermocilla’s study in which they created a Facebook app to crowdsource data in creating a geographical information system that shows the affected areas of disasters in Laguna [21]. Llaguno also created a Facebook app in his study to access the status updates of users for sentiment analysis which can used by advertisers, movie staffs and other organizations that would like to know the reactions of their customers towards their products [13].

Aside from Facebook, there are also studies that used Twitter for crowdsourcing or as a source of data for solving environmental or health-related issues. An example of which is Aramaki et al.’s study in which they detect influenza epidemic by using Support Vector Machines (SVM) classifier on the collected tweets throughout a certain period of time [2]. Another study was by Sakaki et al. in which they predict earthquake in Japan by performing semantic analysis on real-time tweets [19]. Moreover, there is also a study done by Zin et al. in which they used knowledge-based approach to data from Youtube and Twitter for situation awareness in disasters specified such as earthquake and tsunami [27]. Lastly, Osborne et al. also used Twitter and Wikipedia for event detection via first story detection [15].

However, the most relevant studies to this paper are only those that involve the use of social media during disaster, specifically the use of Twitter. This is because most tweets posted during emergency events such as disaster are more about relaying information [10]. Furthermore, it is discovered that during a natural disaster, a significant portion of the tweets gathered promoted situational awareness [26].

2.2 Machine Learning Algorithms

2.2.1 Naïve Bayes

Naïve Bayes classification is a probabilistic learning method. Basically, it computes the probability that a document belongs to a specific class by checking how much evidence is contributed by the terms found in the document [14]. In other words, if applied to this study, the number of disaster-related keywords in a tweet and their relevance will be checked to determine if it’s a disaster-related participant tweet. Some of the advantages of using Naïve Bayes classification are its simplicity, computational efficiency, and good classification performance especially if the data set is large. Its good classification performance can be proved through Llaguno’s study in which he used Naïve Bayes, Rocchio Classifier and Averaged Perceptron to classify whether a particular status update is a positive or negative sentiment [13]. Even though the corpus consisted of only 2,260 status updates, his study showed that out of the three machine learning algorithm that he had used, it’s Naïve Bayes that gives the highest value of accuracy, recall and precision with the value of 0.75, 0.76 and 0.71 respectively. Moreover, Naïve Bayes is also fast and easy to implement which is the basically the reason why it is often used as a baseline in text classification [16].

However, Naïve Bayes also has disadvantages if used in text classification. One of which, according to Rennie et al., is that it assumes that features are independent which yields to a case in which although the words are dependent, each of the word contributes evidence individually [16]. They also pointed out another problem with Naïve Bayes which is it selects poor weights for the decision boundary if one class has more training examples than another. They reasoned out that this is due to Naïve Baye’s under-studied bias effect that shrinks weights for classes with few training examples. Moreover, Corani and Zaffalon showed in their study another disadvantage of Naïve Bayes which is it is too optimistic in dealing with small data sets and missing data which in turn yields to unreliable predictions [6]. This problem can clearly be seen in Llaguno’s study in which using Naïve Bayes in classifying a status update into a positive or negative sentiment only gives a 0.75 accuracy since the corpus only consisted of 2,260 status updates [13]. Moreover, they also showed that Naïve Bayes is also unreliable when it comes to instances that are hard to classify or when computing for posterior probabilities.

2.2.2 Support Vector Machines

Support Vector Machines (SVM) is a large-margin classifier which means it is a vector space based machine learning method where the goal is to find a decision boundary between two classes that is maximally far from any point in the training [11]. Its classification is based on which side of the boundary an instance falls on, given that that instance is mapped into that same vector space. In other words, when applied to this study, the distance of the contents of a tweet from the decision boundary determines whether that tweet is a participant tweet or not.

One of the advantages of SVM is its good generalization capacity in small-size training set problem with high-dimensional input space [5]. Joachims supported this claim through his study in which he compared the performance of SVM with conventional learning methods [11]. In his study, he used 1000 different and relevant features and found out that out of the machine learning algorithms he had used, SVM yields the highest accuracy. He asserted that the reason for this is that SVM avoids the problem of having high dimensional input space. He further asserted that it's due to SVM's use of overfitting protection which does not necessarily depend on the number of features which also means that SVM can handle both large and small feature spaces. Another advantage of SVM that he pointed out is that SVM takes advantage of the fact that most text classification or categorization problems are linearly separable since its main idea is to find linear or polynomial separator/s.

To further prove the good classification SVM gives, Aramaki et al. showed it in their study of detecting influenza epidemic via Twitter [2]. In their study, they have used SVM based sentence classifier in extracting the positive influenza tweets from the gathered tweets and filter out the negative influenza tweets with a f-measure of 0.76. With SVM's accuracy in filtering out negative influenza tweets, their proposed method of detecting influenza epidemic yielded a high accuracy where correlation ratio is 0.89 which outperforms the query-based approach of Google.

Another study which makes use of SVM's advantage of being able to efficiently classify even with highly dimensional data set is by Dilrukshi et al. [8]. In their study, they have used SVM in classifying headlines tweets from Sri Lankan Twitter news groups into 12 specific groups and these groups were chosen in order to cover the main areas of a general news provider. Basically, the aim of their study is to help users identify the most popular news group in Sri Lanka so that they can get brief information about the current state of the country by providing necessary information about development, war, and education among others. By using SVM provided with 90% of their data as training set, they were able to get good classification results for Entertainment, Health, Education and Economy-business groups with more than 75% accuracy.

However, using SVM for text classification also has problems. One of which is that the computational and storage complexity of training is quadratic in the size of the training set. Moreover, its worst case scenario is that it can degenerate to its nearest neighbor, which means that every training point is a support vector, and is much slower to train. Lastly, it has no direct multi-class formulation.

2.2.3 Other Machine Learning Algorithms

SVM active learning is a machine learning algorithm proposed by Tong and Koller [22]. It is a new algorithm for choosing which instances to request next in active learning. Initially, SVMs are generally applied using randomly selected training set classified in advance, and Lewis and Gale (as cited in [22]) proposed a solution to this problem by introducing the idea of pool-based active learning in which the learner has access to a pool of unlabeled data and can request the true class label for a certain number of instances in the pool. However, this algorithm has an issue which is to find a good way of choosing good requests from the pool. This is what Tong and Koller is trying to solve in their study by proposing that in learning, one should choose the best next unlabeled instance to be queried next that best gives information [22]. For this algorithm, they experimented with three approaches: Simple Margin, MaxMin Margin and Ratio Margin. After experimentation, they concluded that if asking each query is expensive relative to computing time, MaxMin or Ratio Margin should be used. On the other hand, if the cost of asking each query is relatively cheap and more emphasis is placed upon fast feedback, then Simple Margin should be used instead. Moreover, they also showed that a hybrid of simple and ratio margin also yields good results. Lastly, they showed that with their algorithm, the need for large labeled training set is reduced and the need for manually labeling the training set is also removed.

There also exist another machine learning algorithm called Naïve Credal Classifier 2 (NCC 2) which was proposed by Corani and Zaffalon [6]. The algorithm is an improvement of Naïve Credal Classifier (NCC) which is a set-valued counterpart of Naïve Bayes. With this algorithm, they extended NCC to a very general and flexible treatment of incomplete data, both in learning and testing. Through this algorithm, they want to solve the problems with Naïve Bayes discussed earlier. Moreover, this algorithm's focus is on pattern classification and its development is based on conservative inference rule (CIR) to compute (imprecise) conditional expectations with incomplete data. Basically, the edge of this algorithm over other traditional classifiers is that it makes generalized set-valued classifications which mean that it issues a determinate classification only when it deems that there's enough information to do so.

In order to show that SVM's text classification performance is better compared to other machine learning approaches, Table 1 and Table 2 are shown. Table 1 shows the performance value in terms of accuracy while Table 2 is in terms of f-measure (F_1 score).

Table 1. Performance of Machine Learning Approaches (Accuracy)

Machine Learning Approach	Performance (Accuracy)
Averaged Perceptron	0.46 (Llaguno, 2013)
Multinomial Naïve Bayes	0.58 (Rennie et al., 2003)
Naïve Bayes	0.75 (Llaguno, 2013)
Rocchio Classifier	0.73 (Llaguno, 2013)
SVM (linear)	**0.93 (Rennie et al., 2003)**
Transformed Weight-normalized Complement Naïve Bayes	0.92 (Rennie et al., 2003)

Table 2. Performance of Machine Learning Approaches (f-measure)

Machine Learning Approach	Performance (f-measure)
AdaBoost	0.592 (Aramaki et al., 2011)
Bagging	0.739 (Aramaki et al., 2011)
Decision Tree	0.698 (Aramaki et al., 2011)
Logistic Regression	0.729 (Aramaki et al., 2011)
Naïve Bayes	0.741 (Aramaki et al., 2011)
Nearest Neighbor	0.695 (Aramaki et al., 2011)
Random Forest	0.729 (Aramaki et al., 2011)
SVM	**0.947 (Dilrukshi et al., 2013)**
SVM (linear)	0.737 (Sakaki et al., 2010)
SVM (polynomial)	0.756 (Aramaki et al., 2011)
SVM (RBF)	0.738 (Aramaki et al., 2011)

As shown in Table 1 and Table 2, it is SVM that yields the highest values for accuracy and f-measure. The values indicate that SVM is indeed a better machine learning approach compared to others. However, it must be noted that some of the values came from a different study which could have different experimental set up.

From these related literatures, the proponents found out that it is in indeed possible to identify disaster-related participant tweets in Twitter and arrive at the decision to specifically use SVM in the tweet classification phase.

3. METHODOLOGY

3.1 Research Design

In this study, the proponents employed text mining and text classification. Text mining, which is also sometimes called as text data mining or simply data mining, is a process of extracting high-quality information from text [20]. In application to this study, the proponents performed text mining in the tweets of Twitter users with the aim of identifying the participant tweets of disaster-related tweets from Philippines. Consequently, the proponents also employed text classification which is basically a problem of assigning a document to one or more classes or categories [7]. In application to this study, the proponents employed text classification on the tweets to determine if they are participant tweets or not. Moreover, the proponents also used text classification on the identified participant tweets in order to group them into three types of disasters which the participant may got affected to: a) flood, b) earthquake, and c) others (fire, landslide, etc).

3.2 Research Procedure

In this section, the research procedure employed in the study will be discussed. Specifically, the research procedure employed in this study involves data gathering, class labeling, data cleaning, data representation and lastly the SVM classifier which will identify if a tweet is a disaster-related participant tweet and will further identify what type of disaster the participant is experiencing. Figure 1 shows the summary of the research procedure employed in the study.

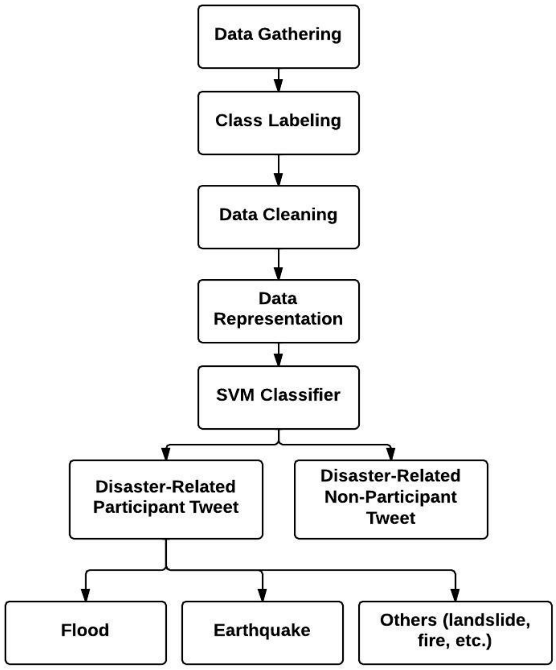

Figure 1. Research Procedure

3.2.1 Data Gathering

In this study, the proponents both used Twitter Search API and Twitter Streaming in gathering public tweets from Twitter. In the initial stages of the study, the proponents were not able to set-up the Twitter Streaming API so the Twitter Search API was used instead in gathering flood-related and earthquake-related tweets. The flood-related tweets is composed of tweets gathered on the onset of Super Typhoon Haiyan (Yolanda) which was just before the power was cut and a portion (15,000) of the procured tweets on 2012 flooding in Luzon caused by Southwest Monsoon (Habagat). In gathering the flood-related tweets, the following keywords and hashtag were used: ulan, bagyo, baha, flood, typhoon, #YolandaPH, #ulanph, #rescueph, #PrayforPhilippines, and #BangonPH. Meanwhile, earthquake-related tweets were gathered on the week the 7.2 magnitude earthquake occurred in Central Visayas by using the following keywords and hahtags: linog, earthquake, shaking, aftershock, landslide, tremor, lindol, quake, #PrayforVisayas, #PrayforBohol, and #rescueph. Lastly, fire-related tweets were gathered from March 6, 2014 to March 16, 2014 through Twitter Streaming API by using the following keywords: fire, apoy, sunog, burn, flame, burn and kalayo. From this, the proponents were able to get around 7,000 flood-related Yolanda tweets, 10,000 earthquake-related tweets, and 30 fire-related tweets.

3.2.2 Class Labeling

In class labeling, a disaster-related tweet can either be labeled as a participant tweet or a non-participant tweet. The term “participant” tweet was first encountered by the proponents in a study by Lee et al. [12]. In their study, a tweet is labeled “participant” if it expresses a user’s first-hand experience of the flooding, or other consequences of it such as being stuck in traffic. Generally, a participant tweet is a tweet by a Twitter user who experienced a disaster first-hand.

Using their study as reference together with the works of Sakaki et al. [19] and Aramaki et al. [2], the proponents were able to create rules in labeling a disaster-related tweet a "participant" tweet. Specifically, the rules are the following:

1. There should be a mention of a place where the disaster happened and it may not be a name of a specific place or location. Mentions of "in the city" or "in the neighborhood" or even a simple "here" or "there" as a place where the disaster occur may suffice. If no location or place is specified, then it will be assumed that the disaster occurred in or near the user's current location.
2. The tweet should be an affirmative statement and not an interrogative or subjunctive statement. For example, a tweet like "Is there an earthquake last night?" will not be considered as an earthquake-related disaster tweet for it still needs further confirmation.
3. The tweet should not be general news or an advisory from Twitter accounts of news groups or government agencies since their tweets is already a form of information dissemination of first-hand experiences of Twitter users of the disaster. In short, their tweets are not their first-hand experience of the disaster.
4. The tweet should not be purely a retweet of a disaster-related tweet because retweeting a tweet does not necessarily mean that the "retweeter" is experiencing the same thing with was indicated in the tweet. During a disaster, more often than not, a Twitter user purely retweets a disaster-related tweet just to help in the information dissemination of it.

If a disaster-related tweet does not satisfy these rules, it is automatically labeled as a non-participant tweet.

As an example, the following flood-related tweets were labeled as participant tweets:

"the flood's inside na!!!!!!!! tae."
"the flood (water) is inside already! shit"
"woooh sana tumigil ulan para bumaba yung tubig sa sala"
"woooh I hope the rain stops so that the (flood) water level in the sala will sink"

The first tweet is a flood-related participant tweet since the Twitter user indicated that the flood water is already inside of where it is staying. The second tweet is a flood-related participant tweet since the Twitter user indicated that the (flood) water is already inside the sala and hoped that the rain would stop so that the (flood) water level will sink.

On the other hand, the following flood-related tweets were labeled as non-participant tweets:

"@_bojaaaa tama :)) baha din ba sa inyo?"
"@_bojaaa correct! :)) Is it also flooding at your home?"
"kung mag baha dre baaa pwede adto ta tog sa hotel?"
"if we'll be flooded here, can we just stay at the hotel?"

The first tweet is a flood-related non-participant tweet since it contains an interrogative statement. Although the keyword "din" may suggest that the Twitter user is already experiencing the flood, it is not conclusive since it may also refer to other person whom both the Twitter user and the Twitter user he/she mentioned may know. The second tweet is a flood-related non-participant tweet since it contains a conditional statement in which the Twitter would like to sleep at a hotel if the flood would come to his/her vicinity. The tweet means that the Twitter user have not experienced the flood yet.

Lastly, the class labeling was performed semi-automatically by creating a program that filters out purely retweeted tweets which start with "RT", interrogative tweets which contain "?" and tweets that contain URLs which are assumed to be from news groups. These filtered tweets were then automatically labeled as non-participant tweets. Moreover, the program also filtered out the participant tweets by using the past and present tense of keywords for each disaster type which are "baha", "flood", "linog", "lindol", "shake", "earthquake", "tremor", "aftershock", "sunog", "fire", "flame", "sunog", "burn" and "kalayo". These filtered tweets were then initially labeled as participant tweets. The rest were then manually labeled by the proponents and the proponents also double-checked the disaster-related tweets initially labeled as participant tweets by the program.

3.2.3 Data Cleaning

In data cleaning, the proponents used the work of Dilrukshi et al. [8] as reference. Specifically, in data cleaning, the proponents first converted the string tweets into lower case for easier comparison and then removed the stop-words (English, Tagalog, & Visayan), emoticons, RTs, mentions (@mention), expressions and other special/miscellaneous characters present in each tweet. The proponents only removed the '#' in #hashtags and not the hashtag itself since there are tweets where each word is a hashtag but actually a sentence when read. Moreover, the proponents' list of English stop-words[1] contains 570 words, the Tagalog contains 123 stop-words and the Visayan contains 110 stop-words. The list of Tagalog and Visayan stop-words were both created by the proponents and were basically a compilation of the Tagalog and Visayan articles, linking verbs, connectives, pronouns and conjunctions.

After removing the stop-words and other miscellaneous characters, the proponents then employed the bag-of-words model [3] and created a pool or dictionary of unique words present in the collection of tweets for each disaster type. In other words, all the words that remained after stop-words and miscellaneous characters removal were used as features. Specifically, the dictionary for flood-related tweets data set contains 7,312 unique words, 9,237 unique words for earthquake-related tweets data set and 12,202 unique words for disaster-related tweets data set. Lastly, the words in the dictionary for each disaster type were sorted alphabetically to facilitate easier index mapping that will be later implemented in data representation.

3.2.4 Data Representation

In data representation, the proponents created string vectors by computing the tf-idf of each word in a tweet as the feature value and the index being the word's index in the dictionary. The proponents used the concept of tf-idf for data representation instead of the simple boolean presence or absence of a word because tf-idf is much informative. tf-idf is much informative because it considers the relative frequency of a word in a tweet to its frequency in the whole collection of tweets in the data set. Specifically, term Frequency (TF) of a word in a tweet was computed by counting the number of occurrences of a word in a tweet and dividing it by the total number of words in the tweet.

---

[1] Source: http://jmlr.org/papers/volume5/lewis04a/a11-smart-stop-list/english.stop

On the other hand, Inverse Document Frequency (IDF) of a word was computed by counting the total number of tweets collected, dividing it the by the number of tweets where the word appears and computing the logarithm of the quotient in base 10. Consequently, TF-IDF of a word was then computed by multiplying the TF and IDF values of a word in a tweet. Figure 2 and Figure 3 show the summarized formula used in computing the tf and idf in the study.

$$\text{tf} = \frac{\text{frequency of a word in a tweet}}{\text{total number of words in a tweet}}$$

Figure 2. tf formula

$$\text{idf} = \log\left(\frac{\text{total number of tweets}}{\text{frequency of a word in the collection of tweets}}\right)$$

Figure 3. idf formula

When a word has a high TF-IDF value, it means that the word is relatively relevant to the collection of tweets. On the other hand, when a word has a low TF-IDF value, it means that the word is possibly just a noise or a stop-word. Figure 4 shows a sample of string vectors created from tweets, with ":" as separator of the word's index in the "dictionary" and of its TF-IDF value.

```
101:0.3085431 446:0.3636406 814:0.34153634 815:0.45472562
1:0.139794 162:0.33064252 808:0.60587674 809:0.5706179
```

Figure 4. Sting Vectors Sample

For the training and testing set, +1 or -1 was inserted as prefix in each string vector to signify that it belongs to the positive class (participant tweet) or negative class (non-participant tweet).

3.2.5 SVM Classifier

In this study, Support Vector Machines (SVM), a supervised machine learning algorithm, was used because it provides better results than other machine learning algorithms, specifically Naïve Bayes, when it comes to sparse or high-dimensional data set [11]. Moreover, SVM has overfitting protection, do not response for local minimum and has the ability to find the global minimum [8].

In implementing the SVM classifier, Chang and Lin's [4] libSVM software was used. In this study, a disaster-related tweet will be identified if it is a participant tweet or not. Furthermore, if it turned out to be a "participant tweet", it will be further identified what type of disaster that participant is experiencing which can be flood, earthquake or others (fire, landslide, etc). Thus, this system's SVM classifier via libSVM was trained with data sets thrice: disaster-related tweets data set first, then flood-related tweets data set and lastly with earthquake-related tweets data set. Moreover, each of these data sets contains equal number of participant and non-participant tweets. Lastly, the SVM classifier was implemented as a binary classification for each disaster type with RBF kernel.

4. RESULTS AND DISCUSSION

In this section, the proponents will discuss the results of the experiments performed in the study. Specifically, the classification performance of the SVM classifier in identifying the disaster-related participant tweets and the disaster type it belongs to will be discussed.

In this study, the proponents had set up an experiment to classify a disaster-related tweet whether it is a participant tweet or a non-participant tweet and identify the disaster type it belongs to by using SVM classifier. Figure 5 shows the general experimental set up used by the proponents for the disaster-related tweets and on each of its type.

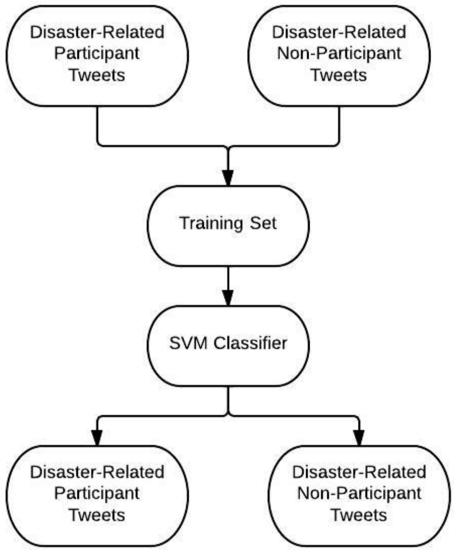

Figure 5. Experimental Set up

As shown in Figure 5, the proponents trained the SVM classifier with the prepared training set which consists of equal number of disaster-related participant tweets and disaster-related non-participant tweets. Actually, the SVM classifier was supposed to be trained first to identify disaster-related tweets from disaster-unrelated tweets but the proponents did not include this part. The proponents skipped this part of training since during data gathering, a disaster is already happening, hence, almost all of the tweets gathered are disaster-related. As a result, the proponents proceeded in training first the classifier to identify disaster-related participant tweets from disaster-related non-participant tweets. Next, if a given tweet or set of tweets are identified as disaster-related participant tweets, the SVM classifier will be retrained with training set that contains participant and non-participant tweets of the disaster types specified until it is identified that is a participant tweet of "others" group (fire, landslide, etc). Specifically, it will be trained with flood-related tweets data set first then followed by the earthquake-related tweets data set. In other words, a multi-level classification was performed in this study and Figure 6 shows the summary of this.

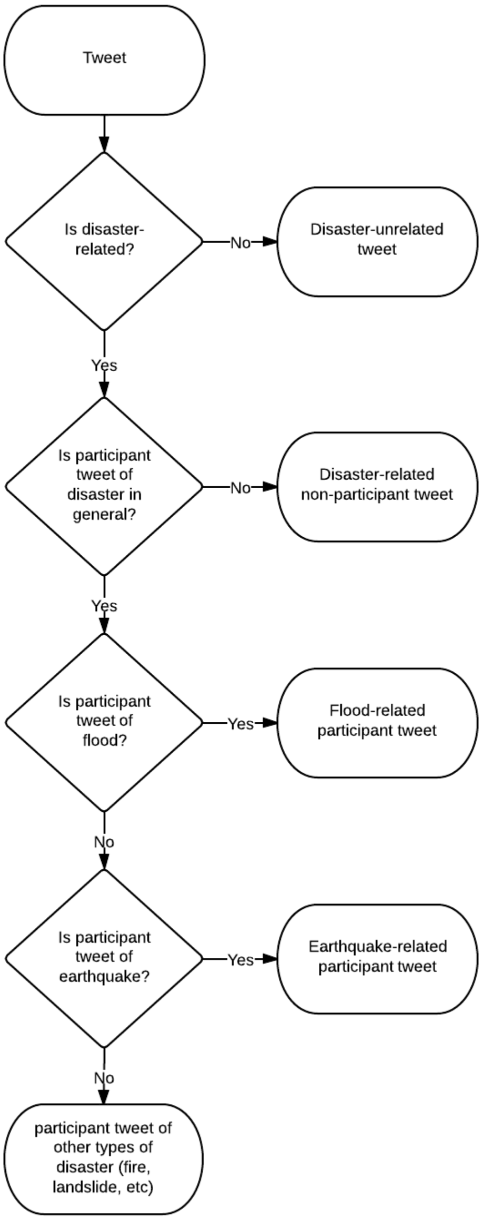

Fig. 6. Flowchart of Multi-level Classification

4.1.1 10-fold Cross-Validation

The proponents performed 10-fold cross-validation on each data set to validate the performance of the SVM classifier and also to select the best model to be used for tweet classification for each data set [13], [8]. In doing so, the classification of SVM classifier will not be biased and limited to a specific set of tweets. Specifically, 80% of the tweets from each data set were chosen randomly for cross-validation set and another 20% were chosen randomly for the final testing set. The disaster-related tweets data set contains 4,286 tweets for each class (participant and non-participant), 2,135 tweets for each class in flood-related tweets data set and 3,294 tweets for each class in earthquake-related tweets data set. The proponents purposely made the data set to contain equal number of participant and non-participant tweets in order to avoid biased classification to either of the class (participant or non-participant). Moreover, while randomly selecting tweets for cross-validation, the proponents also made sure that there would be no duplicates.

For 10-fold cross-validation, the proponents computed the precision, recall and F_1 score of each model. The precision was computed in order to know the exactness of the SVM classifier if that model will be used. Moreover, recall was computed in order to know the completeness or sensitivity of the SVM classifier by using that model. Lastly, the F_1 score was computed in order to have a single valued measurement or basis in selecting the best model and also to give equal weights or to achieve tradeoffs between precision and recall. Table 3 shows the 10-fold cross-validation results of disaster-related tweets data set, Table 4 for the flood-related tweets data set and Table 5 for the earthquake-related tweets data set.

Table 3. Disaster-Related Tweets 10-fold CV Results

Model #	Accuracy	Precision	Recall	F_1 Score
1	57.75	72.88	25.15	0.3739
2	62.87	73.54	40.65	0.5235
3	61.26	70.97	38.60	0.50
4	58.33	68.59	31.29	0.4297
5	64.77	73.39	46.78	0.5714
6	61.40	71.91	37.43	0.4923
7	58.33	71.64	28.07	0.4034
8	**70.32**	**70.71**	**69.88**	**0.7029**
9	65.20	72.73	49.12	0.5864
10	60.23	69.02	37.13	0.4829

Table 4. Flood-Related Tweets 10-fold CV Results

Model #	Accuracy	Precision	Recall	F_1 Score
1	69.71	63.49	94.11	0.7583
2	72.06	65.47	95.88	0.7780
3	75.00	67.62	97.06	0.7971
4	68.53	62.50	94.12	0.7512
5	71.76	65.18	94.71	0.7722
6	68.82	62.31	95.29	0.7535
7	**77.94**	**73.17**	**88.24**	**0.80**
8	73.53	66.81	93.53	0.7794
9	72.65	65.85	95.29	0.7788
10	72.94	65.60	96.47	0.7810

Table 5. Earthquake-Related Tweets 10-fold CV Results

Model #	Accuracy	Precision	Recall	F_1 Score
1	63.69	58.71	93.54	0.7214
2	65.40	63.49	73.38	0.6808
3	64.45	59.18	93.16	0.7238
4	**64.64**	**59.56**	**93.54**	**0.7278**
5	64.26	58.91	94.30	0.7251
6	65.02	60.05	90.87	0.7231
7	64.07	60.22	82.89	0.6976
8	64.26	59.84	86.69	0.7081
9	64.26	59.50	90.49	0.7179
10	64.26	60.10	87.07	0.7112

As shown in Table 3, out of the models for disaster-related tweets data set, it is Model # 2 that gives the highest value of F_1 score which was 0.7909 and thus was chosen as the best model for identifying participant tweets in the disaster-related tweets data set since F_1 score already considered the values for precision and recall. Moreover, as shown in Table 4, the best model for identifying participant tweets in the flood-related tweets data set is Model #7 with the F_1 score of 0.80. Lastly, as shown in Table 5, Model # 4 yields the highest value for F_1 score which is 0.7278 and thus was chosen the best model in identifying participant tweets in the earthquake-related tweets data set. These models were then used in classifying the final testing set for each data set. Table 6 shows the classification performance of these models on the final testing sets.

Table 6. Final Testing Results

Final Testing Set	Accuracy	Precision	Recall	F_1 Score
disaster-related	72.64	71.49	75.50	0.7344
flood-related	81.26	75.14	93.44	0.8330
earthquake-related	64.29	59.26	91.95	0.7207

Table 6 shows the classification performance of the models of each data set. For the disaster-related tweets data set, the model has the following performance: 72.64 accuracy, 71.49 precision, 75.50 recall and 0.7344 F_1 score. This means that the model for the disaster-related tweets data set is well trained in identifying participant tweets than on identifying non-participant tweets. Moreover, since the threshold or accuracy is not that high, the precision is also not that high and consequently, the recall is high. In other words, in every classification of a participant tweet in disaster-related tweets data set, the model is 72.64% confident that its classification correct but is actually just 71.49% correct and since the model is not that confident, a greater percentage of the data set/collection, specifically 75.50%, are classified as so. . Thus, it can be inferred that if the accuracy was high, then surely the precision will increase but the recall will decrease. Lastly, the 0.7344 F_1 score value means that the model's precision and recall values are quite balanced since in order to have high F_1 score value the model should also have high values for both precision and recall.

On the other hand, Table 6 shows that in the flood-related tweets data set, the model has an accuracy of 81.26 which leads to having a lower precision compared to its recall. Specifically, the model has a precision of 75.14 and recall of 93.44. This means that the model for flood-related tweets data set was able to classify 93.44% of the data set and 75.14% of these classifications are correct. These values for precision and recall could also mean that the model is well trained in identifying flood-related participant tweets than on the flood-related non-participant tweets. Lastly, the model has an F_1 score of 0.8330 which just means that both precision and recall of the model are fairly high.

Lastly, Table 6 shows that the model for earthquake-related tweets data set has an accuracy of 64.29, precision of 59.26, recall of 91.95 and F_1 score 0.7207. These values mean that with the accuracy/threshold of 64.29, the model is 59.26% in its classification which leads to a having great percentage (91.95%) of the data set being classified as so. The value for F_1 score just basically means that the model's values for both precision and recall are not that high and balanced.

However, out of the three models, it is the model for flood-related tweets data set that has the highest value for F_1 score which means that this model's precision and recall values are well balanced compared to the models of other data sets. Moreover, this could also mean that this system's SVM classifier can identify well participant tweets of flood than on disaster itself or earthquake.

4.1.2 Error Analysis

In the error analysis, the cause of the low accuracy value and precision value of the classifier in each data set was the false positive classifications made.

For example, the following flood-related non-participant tweets were classified as participant:

"RT @matamaanka: Needs help! 3 senior citizens. 113 Brgy. Sta Teresita"
"A cathedral in Leyte ripping apart piece by piece. A man carrying his dead daughter. The rescue team plucking casualties from flood water."

Note that both tweets are supposed to be classified as non-participant tweets because the first tweet was just asking for help, there is no mention of the flood and it is even purely a retweet while the second tweet is a general description of what the user has observed from the effects of disaster. However, after removing the stop-words and other miscellaneous words in the tweets, the resulting texts are the following:

"senior citizens brgy sta teresita"
"cathedral leyte ripping piece carrying dead daughter rescue team plucking casualties flood water"

It can be noticed that after processing the first tweet, the words "needs" and "help" which are not supposed to be stop-words were removed. The reason for this is that those words are included in the list of English stop-words[2] used by this study. Thus, it can be inferred that one of the possible reasons of false positive classifications done by the SVM classifier is that some words such as negation words or words used in subjunctive or interrogative tweets which could have clearly described a non-participant tweet and differentiated it from a participant tweet may have been removed in the process of stop-words removal. As a result, a supposedly non-participant tweet becomes a participant tweet after stop-words removal.

Moreover, the specified tweets were wrongly classified as participant tweets because of the low tf-idf value of the words present in each tweet after being processed. Since most of the words present in the tweets have low tf-idf value, it means that the words are not relevant enough for the said tweets to be classified as non-participant tweets. Specifically, the words with low tf-idf value in the first tweet are "brgy" and "sta" while for the second tweet are all the words except "piece". As a result, the SVM classifier assumed and classified them as participant tweets. More or less, these reasons also applies to the misclassifications in other data sets since each data set were prepared in the same manner.

4.1.3 System Evaluation

The system assumes that the data that will be fed are disaster-related tweets. Thus, in order to evaluate the system's SVM classifier performance, the proponents selected the first 500 tweets after the fifteen-thousandth mark from Habagat data set since the first 15,000 tweets were already used in creating the

[2] Source: http://jmlr.org/papers/volume5/lewis04a/a11-smart-stop-list/english.stop

flood-related tweets data set. After feeding these 500 tweets into the system, the system returned the following results: 162 out of 500 Habagat tweets were classified as disaster-related participant tweets, and out of these 162 disaster-related participant tweets, 159 were classified as participant tweets of flood-related tweets, and 3 were classified as participant tweets of earthquake-related tweets.

However, the proponents' manual classification returned the following results: out of the 500 Habagat tweets, only 62 were classified as disaster-related participant tweets and all of these are flood-related participant tweets.

The proponents attribute this poor performance to the fact that the models used for classification were static and the size of the data set is relatively small. Thus, it is possible that some of the Habagat tweets being fed into the system contain words that are not found on the bag-of-words or dictionary of words for each data set in which the model was created from or much specifically, in the data set for flood-related participant tweets. As a result, not all words in a tweet are represented in its string vector equivalent and this is crucial because those words might have described the tweet well that it is a participant or a non-participant tweet. However, since those words are not represented, the SVM classifier will just classify a tweet based on what were represented in string vectors resulting to misclassification.

5. CONCLUSION AND RECOMMENDATION

The results of the study showed that it is indeed possible to identify participant tweets of any type of disaster in Twitter using SVM. Specifically, the proponents were able to achieve an F1 score of 73.44, 83.30 and 72.07 for disaster-related, flood-related and earthquake-related participant tweets respectively. However, the models used for each data set were static and the data set was relatively small. As a result, the system yields a poor classification performance when a tweet to be classified contains words that were not taken into consideration by the training sets and consequently by the models used in the system. Moreover, the system has a limitation of assuming that the tweet that will be classified is already filtered out to be a disaster-related tweet.

The results of the study can be used as a starting point in examining if it is possible to identify the disaster-prone areas in the Philippines using Twitter. To the future researchers, the proponents would like to recommend that time, language change over time and actual location of the disaster should be considered. By considering the time the tweet is posted, it may give a possibility to identify or predict what type of disaster usually occurs in a specific season or month. Moreover, by considering language change over time, a change on how Twitter users express their first-hand experiences of a disaster may also be observed. Lastly, by considering the actual location of a disaster as broadcasted by the Twitter accounts of news and government agencies, the credibility of the tweets of the users who experienced a disaster first-hand may be verified.

6. ACKNOWLEDGMENTS

The proponents would like to thank Prof. John Boaz Lee of Ateneo de Manila University for the Habagat tweets that were partly used in the study.

7. REFERENCES

[1] 2013 Bohol earthquake. n.d. Retrieved Feb. 6, 2014 from http://en.wikipedia.org/wiki/2013_Bohol_earthquake

[2] Aramaki, E., Maskawa, S., and Morita, M. 2011. Twitter Catches The Flu: Detecting Influenza Epidemics using Twitter. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* (Edinburgh, Scotland, UK, July 27 – 31, 2011).

[3] Bag-of-words model. n.d. Retrieved September 20, 2013 from http://en.wikipedia.org/wiki/Bag-of-words_model

[4] Chang, C. C., and Lin, C. J. 2011. LIBSVM : a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*. vol. 2, no. 3 (2011), 27:1-27:27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

[5] Chi, M., Feng, R., and Bruzzone, L. 2008. Classification of hyperspectral remote-sensing data with primal SVM for small-sized training dataset problem. *Advances in Space Research* 41 (February 2008), 1793-1799.

[6] Corani, G., and Zaffalon, M. (2008). Learning Reliable Classifiers From Small or Incomplete Data Sets: The Naive Credal Classifier 2. *Journal of Machine Learning Research*, 9 (April 2008), 581-621.

[7] Document classification. n.d. Retrieved October 14, 2013, from http://en.wikipedia.org/wiki/Document_classification

[8] Dilrukshi, I., de Zoysa, K., and Caldera, A. 2013. Twitter News Classification Using SVM. In *The 8th International Conference on Computer Science & Education* (Colombo, Sri Lanka, April 26-28, 2013).

[9] Gao, H., Barbier, G., and Goolsby, R. 2011. Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Transactions on Electron Devices*. vol. 20, no. 3 (2011), 10–14.

[10] Hughes, A., and Palen, L. 2009. Twitter adoption and use in mass convergence and emergency events. In *Proceedings of the 6th International ISCRAM Conference* (Gothenburg, Sweden, May 2009).

[11] Joachims, T. 1998. *Text Categorization with Support Vector Machines: Learning with Many Relevant Features*. Research Report. University of Dortmund.

[12] Lee, J.B., Ybañez, M., De Leon, M. M., and Estuar, M. R. E. 2013. Understanding the Behavior of Filipino Twitter Users during Disaster. *GSTF International Journal on Computing (JoC)*. vol. 3, no.2 (July 2013).

[13] Llaguno, K. P. 2013. *Facebook Status Sentiment Analysis*. Special Problem. University of the Philippines Cebu.

[14] Manning, C. D., Raghavan, P., and Schutze, H. 2009. *An Introduction to Information Retrieval*. Cambridge University Press. Retrieved from http://nlp.stanford.edu/IR-book/pdf/irbookprint.pdf

[15] Osborne, M., Petrovic, S., McCreadie, R., Macdonald, C., and Ounis, I. 2012. Bieber no more: First Story Detection using Twitter and Wikipedia. In *SIGIR 2012 Workshop on Time-aware Information Access* (Portland, Oregon, USA, August 12 - 16, 2012).

[16] Rennie, J. D. M., Shih, L., Teevan, J., and Karger, D. R. 2003. Tackling the Poor Assumptions of Naive Bayes Text

Classier. In *Proceedings of the Twentieth International Conference on Machine Learning* (Washington DC, 2003)

[17] RP highest in social networking engagement across Asia-Pacific. 2010, April 7. *GMA News Online*. Retrieved from http://www.gmanetwork.com/news/story/187883/scitech/rp-highest-in-social- networking-engagement-across-asia-pacific

[18] Russell, J. 2011, May 15. Philippines named social networking capital of the world. *Asian Correspondent*. Retrieved from http://asiancorrespondent.com/54475/philippines-named-the-social-networking- capital-of-the-world-indonesia-malaysia-amongst-top-10/

[19] Sakaki, T., Okazaki, M., and Matsuo, Y. 2010. Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. In *Proceedings of the 19th International World Wide Web conference* (Raleigh, N. Carolina, Apr. 26–30, 2010).

[20] Text mining. n.d. Retrieved October 14, 2013, from http://en.wikipedia.org/wiki/Text_mining

[21] Tolentino, K. A. S., and Hermocilla, J. A. C. 2012. *Disaster Relief in Laguna: A Geographical Information System Through Crowdsourcing on Facebook*. CMSC 190 Special Problem. University of the Philippines Los Baños.

[22] Tong, S., and Koller, D. 2001. Support Vector Machine Active Learning with Applications to Text Classification. *Journal of Machine Learning Research*, (Nov. 2001), 45-66.

[23] Top 10 social networking sites in the Philippines. n.d. Retrieved Sept. 20, 2013 from http://en.wikipilipinas.org/index.php?title=Top_10_Social_Networking_Sites_in_the_Philippines_(2010)

[24] Typhoon Haiyan. n.d. Retrieved Feb. 6, 2014 from http://en.wikipedia.org/wiki/Typhoon_Haiyan

[25] Typhoons in the Philippines. n.d. Retrieved Dec. 8, 2013 from http://en.wikipedia.org/wiki/Typhoons_in_the_Philippines

[26] Vieweg, S., Hughes, A., Starbird, K., and Palen, L. 2010. Microblogging during two natural hazards events: what Twitter may contribute to situational awareness. In *Proceedings of the ACM conference on Computer Human Interaction* (Atlanta, Georgia, USA, April 10-15, 2010).

[27] Zin, T. T., Tin, P., Hama, H., and Toriu, T. 2013. Knowledge based Social Network Applications to Disaster Event Analysis. In *Proceedings of the International MultiConference of Engineers and Computer Scientists 2013* (Hong Kong, March 13-15, 2013).