CSIRO

Microtext Annotation

Sarvnaz Karimi and Jie (Jessie) Yin

January 2013

Contents

1 Annotation

As microblogs have become a comprehensive repository of real-time information, discovering whether or not a microtext, such as a tweet, contains useful information, and if it does what kind of information it conveys, is important for many social media mining applications. In general, text mining systems based on machine learning techniques require training on examples of human-annotated text. Annotation identifies information of interest as defined by the goal of an application. For example, a social media monitoring tool that watches for customer feedback on a service, is interested in knowing what *people* in certain *places* say about a particular *product* of an *organisation*. All these elements should therefore be identified, which raises the need to annotate them first in small scale for the training purposes.

We are interested in annotating microblog posts (i.e. tweets) that are relevant to specific events: natural disasters such as earthquakes or cyclones, and man-made disasters such as terrorist attacks or riots, that affect a large number of people or community. We developed an annotation scheme to cover three main aspects of each tweet related to these events:

What is the content of the tweet, including whether or not it is related to a specific disaster, and what information about a disaster it contains.

When the *what* aspect of the tweet happened: before a disaster or event, during, or after.

Where does the tweet refer to in terms of geo-locations such as country, city or street name, and map-locatable places such as specific buildings.

In the following sections, we explain our annotation scheme, guidelines for annotators, the data used for annotation, and how this data was collected.

2 Annotation Scheme

An annotation scheme was developed to capture *what*, *when*, and *where* in a tweet. We were particularly interested in first, identifying tweets that are related to natural or man-

For every given tweet, you decide whether or not it is talking about a specific disaster, such as January 2009 bushfire in Marysville, Victoria.

If it IS about a specific disaster, choose what type of disaster is. If the type of disaster is not already listed, choose "other" and name it. If the tweet is ambiguous or you cannot tell what the type of the disaster, choose don't know/ambiguous. Please do not choose don't know/ambiguous if you have already chosen a topic or other. You may choose multiple types of disaster at the same time.

If it is NOT a specific disaster, then you can tell us if it is related to entertainment, advertisement, or other. Generic information about disasters such as "it is flood awareness week" belongs to this category.

Note: All the questions are EXCLUSIVELY on the CONTENT OF THE TWEET.

Figure 1: Guidelines for *What* annotation Part I.

made disasters, and second, estimating the knowledge on disaster impact that is shared in the tweets. In particular, for each tweet we aimed to answer the following questions:

- What is the main topic of the tweet (a natural or man-made disaster, or else, e.g. entertainment, or advertisement);
- Is the tweet personal or news/media sharing;
- When was it tweeted? or in other words, does it contain current information?
- What kind of information does it convey about a disaster? i.e. announcement, help, or damage; and,
- Where did the disastrous event happen?

2.1 What and When in the Tweet

We divided our *what* annotation load into two parts: is the tweet disaster related or not (Part I), and what does it say about a disaster (Part II). For the first part, the guidelines in Figure 1 was provided. The annotation scheme and a snapshot of our annotation interface is shown in Figure 2.

(a) Scheme

Is this tweet talking about a disaster?

Is the tweet related to or anticipating a disaster?
- ◯ Yes
- ◯ No
- ◯ Tweet does not load

If not a specific disaster, what is it about?
- ◯ Entertainment (music/movie/sports/etc)
- ◯ Advertisement/promotion
- ◯ Other

What kind of disaster?
- ☐ Earthquake
- ☐ Flooding (only natural disaster)
- ☐ Fire
- ☐ Storm, cyclone, typhoon, tornado, or hurricane
- ☐ Riot, terrorist attack, or protest
- ☐ Traffic accident
- ☐ Other
- ☐ Don't know/ambiguous

(b) Interface snapshot

Figure 2: "What" annotation: part I. (a) annotation scheme, (b) a snapshot of the interface that shows how tweets were presented to the annotators.

Once annotations for Part I are completed, we proceed with Part II: *what* information is shared about a specific disaster. We were interested in information regarding whether the tweet was about personal experience or media sharing, whether it was about something happening recently or a previous event in the long past, whether it was about asking for help or reporting damage. Information on help and damage is particularly important to identify in emergency situations because they help to decide if a tweet is valuable or not. Therefore, for the second part of annotations, we provided the guidelines in Figure 3, with the annotation scheme shown in Figure 4. Tweets were presented in the same way as Part I.

You will see short messages (tweets) that are related to a specific natural or a man-made disaster. For every given tweet, you will need to answer three main questions:

1. Specify if the tweet is about a personal experience or it is a piece of news or media that someone/media is sharing with others;
2. Specify if the tweet was written before an event, during an event, or after an event.
3. Specify if the tweet announces a disastrous event, requests for help, offers help, or reports damage. For help and damage you will specify the type. An example for announce is: just felt an #earthquake! #melbourne

Note: All the questions are EXCLUSIVELY on the CONTENT OF THE TWEET.

Figure 3: Guidelines for *What* annotation, Part II.

2.2 Where in the Tweet

Location information is crucial for making an association between events and where they took place. It is particularly important in disaster-related messages as they can show emergency responders where issues occurred or where help was needed. For example, if one is to estimate the impact of a storm on an area, affected places are crucial in estimating the scale of the storm. To learn how to automatically identify points of interest (POIs) in microtexts, we decided to first collect human annotations on a set of sampled tweets from a variety of disastrous events. A POI can be any location that is identifiable by latitude and longitude in a map, such as a city, a river, or a shopping centre. POIs can have a variety of granularity ranging from very high-levels such as a country name to low-levels such as a building name.

To identify the "*where*" in the tweets, we provided the guidelines in Figure 5. Annotators were then presented with a tweet where they could choose the location words from a tokenised, stopped list of the tweet words. To reduce the noise we also filtered out those words that could not be a place name, such as swear words or shortened URLs. An example is shown in Figure 6 where we expect the annotator to choose: Louis Vuit-

Is the tweet repeating or re-sharing news or other media?
- ◯ Yes
- ◯ No

Is the tweet related to before, during or after the disaster?
- ◯ Before
- ◯ During
- ◯ After

Does the tweet
- ☐ Announce a disaster
- ☐ Ask for help
- ☐ Offer/announce help
- ☐ Report Damage
- ☐ Other

What kind of help
- ☐ Medical
- ☐ Financial
- ☐ Physical help (food, shelter, labor, etc.)
- ☐ Information about people (e.g. locating people)
- ☐ Other information

What kind of damage?
- ☐ Financial
- ☐ Emotional
- ☐ Property
- ☐ Death
- ☐ Injury
- ☐ Other (e.g. services)

Figure 4: "What" annotation scheme: Part II.

ton store, CBD, #bnefloods, and #qldfloods as point-of-interest.

3 Data

We define *Twitter data* as the content of short messages (tweets) posted by Twitter users with their corresponding information, including tweet unique identifier (tweet-id), publication date, and author's specifications as user handler or username, real name, and location. We collected *Twitter data* in three different ways:

Keyword search: a keyword search using Twitter search API;

User-specific: a sample of tweets from Twitter accounts of specific users such as emergency services or police;

Category	Search Kewords	# Tweets
Earthquake	earthquake, quake, aftershock quake	24,537
Flooding	flood, flooding, flash flood, river flood, river flooding, flood warning	48,928
Fire	fire, wildfire, bushfire, bush fire, coal seam fire, mine fire, grass fire, grassfire, fire warning	
Cyclone/storm	storm, cyclone, hurricane, typhoon, hurricane, storm, typhoon storm	40,531
Riot/protest	civil disorder, civil unrest, civil strife, riot, sabotage, oppression, protest	35,635
Traffic accident	car accident, traffic accident, motor vehicle collision, motor vehicle accident, automobile accident, road traffic collision, car crash	56,238
Financial crisis	financial crisis, recession, stock market crash, sovereign default, banking crisis, bank run, credit crunch, credit squeeze	76,595
Entertainment	music, song, music fan, pop music, music video, movie, celebrity, entertainment, actor, musician, fashion model, movie director, comedian, television host, TV host, sport, Olympics, game, football, swimming, win, medal, TV, TV show, film, buy, holiday	173,058

Table 1: Tweet categories, search keywords used to capture these tweets, and the number of unique tweets in each category.

Event-specific search: a sample of tweets from the rough time range of specific events that happened in the past were collected from a service called Topsy[1], our ESA database [1], and manual search over Twitter at the time of event occurrence.

te queried Twitter once a day using a large number of search keywords (both plain and hashtags[2]) related to earthquake, storm, cyclone, flooding, fire, riot, protest, civil war, financial crisis, traffic accident, and entertainments such as music, films, TV series, and sports over a period of four months from April 2012 till July 2012. These keywords are listed in the second column of Table 1. A list of tweet categories from this set and the unique number of tweets collected for each category are also shown in Table 1. A total of 455,522 tweets were collected in this set.

A second set of tweets was collected from a set of Twitter users that broadcast on emergency situations such as bushfires in Australia. This set of tweets would be different from other two sets because they are written in a formal language and often they include

[1] http://topsy.com/

[2] Keyword preceded by a #; for example #flood.

Event	Location and Date	Source	# Tweets
Christchurch earthquake	Christchurch, New Zealand (22 February 2011)	ESA	5,849
Melbourne earthquake	Australia (19 June 2012)	ESA	92,382
Cyclone Yasi	Queensland, Australia (3 February 2011)	ESA	6,882
Cyclone Yasi flooding	Queensland, Australia (5 February 2011)	ESA	4,687
Diamant Hotel fire	Canberra, Australia (23 June 2011)	ESA	12922
Black Saturday bushfire	Victoria, Australia (7 February 2009)	Topsy	39
Iranian presidential election	Iran (June 2009)	Topsy	18
Queensland floods	Queensland, Australia (February 2009)	Topsy	129
Other (London riots, Libyan civil war, Mumbai attacks)	World (2008-2011)	Topsy	91
York flooding, QLD storm, Bushfires in NSW, Hurricane Sandy	UK, US, Australia, 2012	Twitter search	3,104

Table 2: Incidents of disastrous events with their date and location of their occurrences, source of finding their tweet-ids, and the number of tweets extracted per event.

more details, especially about locations. We were particularly interested in the location information to help build a set of data that could potentially be used in the training of systems that identify POIs in tweets. We collected a small set of 371 tweets from the following twitter handlers (30 tweets per account on avergae):

- @NSWRFS: NSW RFS is the account for fire services in rural areas of NSW, Australia.
- @CFA_Updates: CFA Updates is official account of CFA (Country Fire Authority) a voluntary and community based fire and emergency serice is Australia.
- @ABCEmergency: ABC Emergency is the account of ABC news that reports on emergency situations in Australia.
- @nswpolice: NSW police is the official Twitter site of New South Wales, Australia Police Force.
- @VictoriaPolice: Victoria police is the official Twitter site of Victoria, Australia Police Force.
- @QLDC_Emergency: Emergency Management provides information on emergencies in the Queenstown Lakes District, Otago Region, New Zealand.
- @BrisbaneFloods: a non-official account that shares information on flooding situations in Brisbane, Australia.

- @QldFire: QldFire is the official account of Queensland, Australia Fire and Rescue Service.
- @LiveTrafficNSW: Live Traffic NSW tweets traffic information, incuding accidents, in New South Wales, Australia.
- @SAPoliceNews: SA Police News is the official Twitter site of South Australia (SA) state Police Force.
- @dfes_wa: DFES or Department of Fire and Emergency Services tweets alerts on fires, for example bushfire, in Western Australia;
- @EmergencyAUS: EmergencyAUS is an official account that shares official and non-official emergency related tweets around Australia.
- @Info4Alerts: Info4Disasters, provides live alerts for disasters around the world.

A set of event-specific tweet-ids were identified from the database of the ESA (Emergency Situation Awareness) [1] project. Using these ids the full tweet content as JSON file were downloaded using Twitter API. From the ESA database, we only used tweets that were published in a short period of time when specific disasters happened, as listed in Table 2. Overall, 122,722 tweets were collected this way. Note that ESA database only contains tweets from Australia and New Zealand.

Topsy search service was also used to search on historical tweets to hand-pick tweet-ids that were related to specific disastrous events (Table 2). Topsy however only provides tweets that were most popular and does not give a good scale on the events. We only extracted tweet-ids from Topsy and then used Twitter API to retrieve the tweet contents directly from Twitter. We collected a small set of 277 unique tweets this way.

We also manually searched on Twitter for tweets related to incidents at the time they were happening. Tweet-ids of these tweets were recorded and later these tweets were downloaded using Twitter API. Table 2 lists the major incidents. This data also includes tweets from flooding in Pakistan and France in 2012, and other tweets off topic (usually only have mentions of location where a disaster happened but were not about that disaster). Total number of tweets collected this way was 3,104.

Preprocessing and Sampling

After removing non-English tweets, duplicates, and retweets, we sampled the data of each category —except for the second set that were from the specific users and the data from Topsy which was small— using a simple strategy. Given official accounts such as news agencies contribute to a large number of tweets related to world events, we decided to penalise their contribution to our sampled data by multiplying a uniform random probability by inverse of total number of tweets as

$$p_t = \frac{1}{n} \times u_t,$$

where u_t is a uniform random probability for tweet t, and n is the number of tweets by the author of the tweet t. Tweets were ranked based on their assigned probability and then top-N tweets were chosen for annotation.

Tweets that belonged to the entertainment and advertisement category were sampled separately and a total of 200 tweets were chosen to be annotated. Given other tweet sets already contain non-disaster tweets, we only included a limited amount of noise from this set.

To sample tweets for *where* annotation we used the pool of tweets that were already considered for *what* annotation. We allocated more weight to tweets that were identified as reporting damage and ask or offer help during a disaster, because these tweets are more likely to contain location information and also it is more useful to find locations in such tweets. We therefore calculated a probability p_t of selecting tweet t for annotation by giving a weight w_t to uniform random probability as below:

$$p_t = w_t \times u_t; \text{ where, } w_t = \begin{cases} 0.5 & t \text{ is damage} \\ 0.3 & t \text{ is help} \\ 0.2 & \text{other} \end{cases}$$

4 Gold Annotation

An initial set of 450 tweets, majority belonging to February 2011 earthquake in Christchurch, New Zealand, were annotated by three annotators[3] to identify the “what” aspects of the

[3] The authors and another researcher from CSIRO.

tweets. During this annotation practice, we adjusted the instructions for each section of annotations (final version is shown in Figure 2) in order to create a clear annotation guideline for the external annotators. We had full agreement on 243 of these tweets (54%), that is all the items in annotation scheme were annotated exactly the same by all the three annotators.

5 Crowd-sourced Annotators

To recruit annotators, we used a crowd-sourcing service called Crowdflower[4] which provides an interface to Amazon Mechanical Turk[5]. To work on our tasks, annotators had to pass a training phase prior to commencing their annotations. Each training phase consisted of four tweets with known annotations. Upon successful completion of the training phase, annotators were presented with actual tasks. A task was annotating a single tweet for a given scheme (*what* or *where*). Tasks were grouped in HIT (Human Intelligence Task) which was five tasks from which one was a gold tweet to constantly test annotators level of trust. We paid one cent (American dollars) per tweet.

We specifically only chose workers from English speaking countries (Unites States, UK, Australia, New Zealand, and South Africa) to be allowed to work on our tasks. Demographics of annotators recruited through Crowdflower were largely biased though with majority (98%) from the United States.

6 Final Collection

A total of 5,747 tweets were reliably –at least two annotators agreed on the annotations— annotated for *what* Part I: is this tweet talking about a disaster and if yes, what kind of disaster. The second set of annotations were done on *what* Part II: what is reported for a disaster. We therefore only annotated a subset of tweets from the first part that was 2,850 tweets. Where annotations were done after the these two stages were completed.

We also annotated the set of 2,850 disaster related tweets for their *where* information. Apart from those tweets annotated as gold, we had 100% agreement between at least two

[4] http://crowdflower.com/solutions/self-service

[5] Other options than Amazon Mechanical Turk was also available but we chose to only use Amazon service.

of of three annotators on 2543 tweets (89%), annotating exact same tokens as location clues.

7 Conclusions

We presented our annotation scheme for Twitter data to be used for extraction of information on disastrous events. We annotated these tweets in different levels, from whether or not a tweet is about a disaster to details it provides about the disaster including location information.

We created a corpus of 5,747 tweets with rich annotations for a variety of disastrous events. We also created a corpus of 2,878 tweets which their point-of-interest information are annotated. Such data is of great value for training and evaluation of systems that intend to automatically discover useful information from large volume of microtext generated daily.

We note that given tweets are very short (140 characters), the content of a tweet is often ambiguous or a tweet itself can not provide sufficient information. Thus, annotating one single tweet for its topic may not always be possible or accurate. In our data, for example, annotators considered 39 of the tweets as ambiguous; even though they could tell they are about some sort of disaster, they can not decide the exact type of disaster.

8 Acknowledgement

We thank David Milne for very useful discussions, help in designing the annotation schemes, and annotating gold tweets. We would like to thank ESA team, Mark Cameron, Bella Robinson, and Robert Power for their support. We also thank Cecile Paris for her constructive comments.

9 Bibliography

[1] Jie Yin, Andrew Lampert, Mark Cameron, Bella Robinson, and Robert Power. Using social media to enhance emergency situation awareness. *IEEE Intelligent Systems*, 27:52–59, 2012.

Your task is for each given tweet, find all the words that indicate or imply a location or place name that one would be able to specifically find in a map (e.g. Google maps). In other words, we are interested in:

- Geolocation: country, city, state, suburb, lane, river, etc.
- Place name: a specific building (e.g. name of a shopping centre or a school), garden, park, organisation, or any place that can be uniquely identified on a map. For example, Early-bird pre-school is a place name but a pre-school is NOT a place name.
- Important notes:
  1. If a location has multiple words, choose all of them. For example, in “there was an accident in Victoria avenue which...” you need to select both Victoria and avenue.
  2. If a location is part of a hashtag, you must choose the hashtag. For example, #eqnz indicates earthquake in New-Zealand. Because it contains a geolocation (NZ), it must be selected.
  3. If a word that implies a location (e.g. Australian) is part of a location name (e.g. Australian Tax Office), it has to be selected, otherwise NOT selected.
  4. Majority of tweets are from Australia, therefore please check unfamiliar place names on Google maps: `https://maps.google.com.au/`. Australian states are: NSW (New South Wales), VIC (Victoria), QLD (Queensland), TAS (Tasmania), WA (Western Australia), SA (South Australia), NT (Northern Territory), ACT (Australian Capital Territory).
  5. If there is NO location, you DO NOT click on anything.

Figure 5: “Where” annotation scheme.

Can you find a location or place name?

Tweet: From handbags to sandbags- Louis Vuitton store in CBD Pic: `http://twitpic.com/3p8fkz` #bnefloods #qldfloods

words

☐ From
☐ handbags
☐ to
☐ sandbags
☐ Louis
☐ Vuitton
☐ store
☐ in
☐ CBD
☐ Pic
☐ #bnefloods
☐ #qldfloods

Unsure? check Google maps: https://maps.google.com.au/ Australian states are: NSW, VIC, QLD, TAS, WA, SA, NT, ACT.

Figure 6: A “where” annotation example.

Contact Us

Phone: 1300 363 400

+61 3 9545 2176

Email: enquiries@csiro.au

Web: www.csiro.au

Your CSIRO

Australia is founding its future on science and innovation. Its national science agency, CSIRO, is a powerhouse of ideas, technologies and skills for building prosperity, growth, health and sustainability. It serves governments, industries, business and communities across the nation.