An Evidence Based Earthquake Detector using Twitter

Bella Robinson
bella.robinson@csiro.au

Robert Power
robert.power@csiro.au

Mark Cameron
mark.cameron@csiro.au

CSIRO Computational Informatics
G.P.O. Box 664
Canberra, ACT 2601, Australia

Abstract

This paper presents a notification system to identify earthquakes from first-hand reports published on Twitter. Tweets from target regions in Australia and New Zealand are checked for earthquake keyword frequency bursts and then processed to identify evidence of an earthquake.

The benefit of our earthquake detector is that it relies on evidence of firsthand 'felt' reports from Twitter, provides an indication of the earthquake intensity and will be the trigger for further classification of Tweets for impact analysis.

We describe how the detector has been incrementally improved, most notably by the introduction of a text classifier. During its initial five months of operation the system has generated 49 notifications of which 29 related to real earthquake events.

1 Introduction

Australia and New Zealand have experienced a number of large scale disaster events in recent years. Christchurch New Zealand suffered two earthquakes of magnitude 7.1 (4 September 2010) and 6.3 (22 February 2011) with significant aftershocks continuing around this time. While there were no reported fatalities for the first event, there were 185 deaths in the second with widespread damage and an estimated NZ$15 billion in reconstruction costs (Bruns and Burgess, 2012).

In Australia, the Victorian 2009 Black Saturday Bushfires killed 173 people, impacted 78 towns with loses estimated at A$2.9 billion (Stephenson et al., 2012). The 2010-2011 floods in Queensland affected 70 towns, including the state capital Brisbane, and caused infrastructure damage of A$8 billion (RBA, 2011). Tropical cyclone Yasi (Feb 2011) was a category 5 system that crossed northern Queensland causing an estimated A$800 million in damage (Qld Budget, 2011).

In order to effectively prepare and respond to emergency situations it is critical that emergency managers and crisis coordinators have relevant and reliable information. In Australia, this knowledge is traditionally obtained from official authoritative sources such as the state emergency services and first responder agencies, such as the police force, fire and rescue and rural fire services. Traditional news media (television, news agency web sites and sometimes radio) is also used to provide intelligence about events.

Social media has been recognised as a potential new source of information for emergency managers (Anderson, 2012; Bruns et al., 2012; Alliance Strategic Research, 2011; Lindsay, 2011; Charlton, 2012). However, it is mostly used 'passively' in that during crisis events the emergency services agencies use social media to disseminate information to the community and receive user feedback on their advice (Lindsay, 2011). In order for the full potential of social media to be realised, it needs to be embraced as a new 'channel' of information where the evolving situational awareness of events can be improved.

This paper presents a case study outlining the use of Twitter to detect earthquakes. The detector generates email notifications summarising the Tweets contributing to the alert and includes an indication of the intensity of the event.

2 Background

2.1 Earthquake Detection

An earthquake results from movement in the Earth's crust and different scales have been defined to measure them. The moment magnitude and Richter scales measure the energy released whereas the Modified Mercalli Intensity (MMI)

scale (Eiby, 1966) measures the effects.

An earthquake in the ocean may produce a tsunami. The tsunamis in Indonesia and Thailand on 26 December 2004 occurred with little warning to the communities affected. The Japan tsunami of 11 March 2011 was preceded by warnings however its size was greater than anticipated and resulted in widespread damage and loss of life.

Tsunami warning centres exist world wide. They rely on the identification of earthquakes and information from networks of sea level monitoring equipment, such as coastal tide gauges and deep ocean tsunami sensors, in conjunction with software models to determine the existence of tsunamis, their intensity and trajectory.

Identifying earthquakes is a time consuming and complex task performed by highly trained seismologists. Verification of an earthquake event requires the use of a global network of seismic stations to determine the precise location and magnitude of the earthquake. This process can take up to 15 minutes from when the earthquake occurred.

While some countries and regions have a highly sophisticated and dense network of seismic sensors, for example Japan[1] has over 4000 sensors and the state of California[2] in the USA has over 3000 sensors, other countries including some that are highly earthquake prone do not. Australia[3] which is roughly 20 times the area of Japan has only 70 sensors, earthquake prone Indonesia[4] is roughly 5 times the size of Japan and has only 400 sensors and New Zealand[5] which is roughly one third smaller than Japan and also earthquake prone has only 300 sensors.

Recent studies (Sakaki et al., 2010; Earle et al., 2012; Sakaki et al., 2013; Robinson et al., 2013) indicate that when an earthquake event occurs in populated regions, reports on Twitter can provide a faster method of detection compared to traditional approaches. The role of seismologists to verify and scientifically characterise earthquakes can be augmented by crowd sourced information that provides both an early warning and evidence of the impact experienced by the community affected.

It is important to note that an earthquake's magnitude and location cannot reliably be used to infer the impact on a community. For example, the first Christchurch earthquake mentioned above caused damage to buildings, but fortunately there was no loss of life. The one that followed was smaller and technically an aftershock of the previous one (Bruns and Burgess, 2012), but resulted in fatalities and extensive damage.

2.2 Related Work

Studies of Twitter communications during crises and natural disasters such as earthquakes, have found strong temporal correlations with real-world events (Mendoza et al., 2010). Applying NLP classifiers to extracting situation awareness information has been investigated by Verma et al. (2011). Again that study finds there is strong correlation between data collected in a localised region about a local event and evidence of situation awareness Tweet content.

Several systems have been developed for the automatic detection of earthquakes via Twitter. Earle et al. (2012) describe a detection system operating over a filtered Tweet stream. Importantly, Tweets are filtered if they contain http, RT or @ symbols or more than n tokens. Their detection algorithm is based on a modified short-term longterm ratio over this filtered stream. The filters and n token heuristic aim to account for non firsthand reports. Sakaki et al. (2010) and Sakaki et al. (2013) have deployed a functional system in Japan that uses natural language processing techniques to classify Tweets containing specific keywords, such as *earthquake* or *shaking*. Positively classified Tweets are then used to generate a probabilistic spatio-temporal model of the event and particle filtering is used to estimate the earthquake location. Users of both systems are notified of a potential earthquake via email.

2.3 Social Media Platform

In order to demonstrate the benefit of information published on social media for emergency management we have been continuously collecting Tweets originating from Australia and New Zealand since March 2010 (Cameron et al., 2012). To date, over one billion Tweets have been processed at approximately 1500 per minute. These Tweets have been used to: experiment with alternative algorithms for event detection; develop clustering techniques for condensing and summarising information con-

---

[1] http://www.jma.go.jp/jma/en/
Activities/earthquake.html
[2] http://www.cisn.org/instr/
[3] http://www.ga.gov.au/earthquakes/
seismicSearch.do
[4] http://aeic.bmg.go.id/aeic/indonesia.
html
[5] http://info.geonet.org.nz/display/
equip/Equipment

tent; develop language models to characterise the expected discourse on Twitter; develop an alerting system based on the language model to detect deviations from the expected discourse; train and evaluate text classification systems; and perform forensic analysis.

The aim is to develop a near-real-time platform that monitors Twitter to identify events and improve the situational awareness of emergency events for emergency managers and crisis coordinators. While originally developed for Australia and New Zealand, the technology can be configured and deployed for any region. Additional work would be required to process languages that do not use spaces to separated words.

3   The Problem

The task is to quickly and reliably detect, locate and estimate the intensity of earthquakes as reported on Twitter. Earthquake detection provides a targeted use case to test our social media platform. While early detection of earthquakes can currently be achieved using traditional methods, for example using seismic equipment, this process is tuned to accurately locate and measure the *magnitude* of an earthquake: the energy released. An important distinction is the earthquake *intensity*: the effect of the earthquake to people and the impact on the natural and built environments.

People Tweeting in response to an earthquake event effectively become sensors indicating the scale of the event in terms of the number of Tweets collected. The Tweet content can also be analysed to provide an indication of impact severity. These measures, the scale in terms of number of Tweets and the severity in terms of Tweet content, can be combined to provide an earthquake intensity measure analogous to the MMI scale.

3.1   Preliminary Work

The Social Media platform described in Section 2.3 was configured using heuristics to identify firsthand 'felt' reports as evidence of earthquake events. The heuristics were arrived at after examining the Tweets for all historical earthquake related alerts reported by the system.

The alerts are generated in reference to a background language model. In essence, a five minute buffer of the most recent Tweets is maintained where the frequency of words in the buffer is compared against an historical model of expected word frequencies. When the observed word frequency deviates significantly from the historical model, an alert is generated. The buffer is advanced in one minute increments thus producing a new set of alerts each minute. These alerts are recorded by the platform in a database.

Note that the alerts generated by this method correspond to bursts of unusual word frequencies with respect to the historical language model, not to bursts in the arrival rate of Tweets.

To determine the heuristics to use, 12 months of historical earthquake alerts were analysed. In summary, the process involves filtering alerts generated from the social media platform that match earthquake related keywords, testing the currency of the alert (only consider the first alert generated and not subsequent ones), determining if the Tweets producing the alert are close geographically and measuring the retweet ratio (since a retweet cannot be a firsthand 'felt' report).

3.2   The Heuristic Detector

An earthquake detector was developed as described above. Heuristic thresholds were identified in reference to the earthquake events recorded in the 12 month analysis period. When an earthquake event is found an email notification is generated summarising the Tweet information and heuristic results.

This email is sent to the Joint Australian Tsunami Warning Centre (JATWC) who have responsibility to detect earthquakes in the oceans around Australia, to identify potential tsunami events when such earthquakes are identified and to issue tsunami warnings as required.

This detector has been in operation since mid December 2012. During the first five months of operation, 49 earthquake emails were generated. These notifications were manually reviewed and 29 found to correspond with real earthquake events (*true positives (TP)*). The remaining 20 (*false positives (FP)*) were a result of discussions about earthquakes but not prompted by an event.

A review after two months of operation identified changes to the thresholds used for the heuristics. Doing so improved the results, greatly reducing the *false positives*, but required extensive work to investigate the detected events by cross referencing with seismically verified earthquakes as listed by New Zealand's GeoNet[5] and Geoscience Australia (GA)[3].

4 Introducing a Classifier

The task of detecting earthquakes from Twitter was then considered as a text classification problem. The results obtained in the first five months of operation described above provided a set of earthquake related Tweets that could be labelled as a test set. These Tweets were used to train a classifier configured using a comprehensive range of features. This process forms the basis of our paper: improving the accuracy of the earthquake detector by incorporating the use of a text classifier to predict whether individual Tweets are instances of firsthand earthquake reports.

The following sections describe the journey we have taken in developing a classifier for earthquake detection.

4.1 Earthquake Alert Annotations

The process of reviewing the performance of the heuristic detector involved examination of the Tweets contributing to each earthquake related alert and labelling them as evidence of firsthand earthquake reports. This produced a collection of 237 alerts, of which 45 contained examples of firsthand earthquake reports. Reviewing the existing heuristic detector's performance in terms of these alerts we achieve an F1 score of 0.667 (TP=23,FP=11,TN=181,FN=12).

Figure 1 shows the number of Tweets that include the word 'earthquake' and contribute to an earthquake alert. These numbers are aggregates of multiple alerts for time periods from mid December 2012. There are 34 such time periods with a total of over 8000 Tweets. 15 of these time periods have 50 or less Tweets in them, 14 have more than 100 and two have more than one thousand.
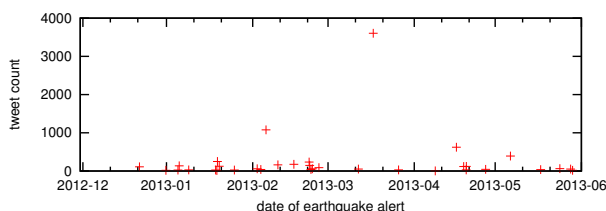

Figure 1: 'earthquake' alerts

The same data is plotted in Figure 2 showing more detail for the time periods with fewer Tweets.

4.2 Initial Training Data

Training data was needed to configure the classifier. Initially, sets of suggested positive and negative Tweets were generated by using the alert
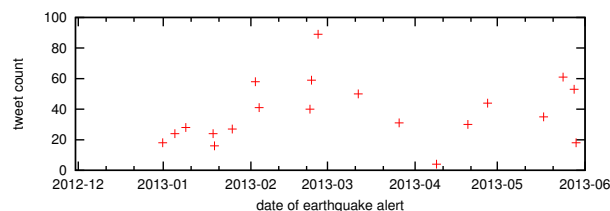

Figure 2: 'earthquake' alerts: zoomed

data identified in the alert annotation step outlined above. All Tweets contributing to an alert labelled as positive were initially labelled as positive and the reverse for the negative alerts. To increase the sample size, Tweets were also gathered from follow-up alerts that occurred within five minutes of the initial alert.

It is important to note that retweets have been excluded from the classification process: by definition a retweet cannot be a firsthand report of feeling an earthquake.

The suggested positive and negative Tweets were then examined individually to adjust the labels when incorrect. The results of this initial Tweet annotation phase produced a set of 1604 labelled Tweets, with 868 being positive and 736 negative. Examples are shown in Table 1.

4.3 Feature Selection

As noted in Joachims (1998), Support Vector Machines (SVMs) are well suited for text categorisation. We have therefore used the LIBSVM (Chang and Lin, 2011) software, configured with the linear kernel function, to perform SVM classification for this work. A number of features can be used to construct a representative vector for each Tweet. For example ngrams (unigrams, bigrams or a combination of both), the number of words in the Tweet, the number of hash tags and hyperlinks used and the number of user mentions.

During the Tweet annotation process a couple of features stood out as being particularly important: firsthand reports are usually short, don't contain a hyperlink and often contain particular words including exclamations. It was unclear whether the other features would be helpful or not.

Note that that all Tweets used in the training process contained either the word 'earthquake' or the hash tag '#eqnz': these are the keywords currently monitored by the heuristic earthquake detector. These particular instances of unigram features should not contribute to the outcome.

To determine which features contribute to the

4

Table 1: Examples of positive and negative Tweets

Firsthand reports	Not firsthand reports
Woah! Earthquake	Magnitude 3.8 earthquake shakes Wellington: Wellingtonians were shaken awake by a magnitude 3.8 earthquake early…http://t.co/rT4UvjzH
Earthquake!! 2 small 3-second-each tremors just now!!	i thought there was an earthquake or some sort of world ending experience but then i realised my brother was running around upstairs.. woops
That was a goodun. #eqnz	Large earthquake struck Vanuatu. Imagine the thoughts running through their heads when the earth started to shake
oooh, big wobble, heard that coming way off #eqnz	Can't believe it's been 2 years today since Christchurch had its major earthquake #KiaKaha

task of identifying firsthand earthquake reports, a 10 fold cross validation process (Hastie et al., 2009) was used with all combinations of the following features: ngram combinations, Tweet length, hash tag count, user mention count and hyperlink count. Table 2 shows a subset of the results, including the accuracy measures achieved by training on each single feature only, plus the highest scoring feature combination.

Table 2: Training Results (averages)

Feature	Accuracy (% correct)	F1 score
Tweet length (word count)	77.9	0.796
User mention count	64.2	0.746
Hash tag count	65.8	0.654
Hyperlink count	73.1	0.800
Unigrams	86.7	0.882
Bigrams	80.7	0.843
Combo of uni and bigrams	86.5	0.880
Unigrams plus all others	90.3	0.912

As expected, the words used within each Tweet (ngrams), the hyperlink count and Tweet length all perform well by themselves, with the user mention count and hash tag count in particular being less important. The combination of all of these features however, produces the highest average accuracy and F1 score.

4.4 More Training Data

The combination that produced the best score in the feature selection process was used to train a classifier using the annotated Tweet data described in Section 4.2. It was not appropriate to use this classifier to revisit the accuracy of the earthquake alerts since the Tweets contributing to each alert were used to train the classifier. Instead, a new training dataset was created from Tweet data before December 2012; before the heuristic detector was deployed.

The classifier was used to aid this process. The Tweets contributing to historical earthquake alerts (pre December 2012) were processed by the classifier to generate roughly 2000 suggested positive and negative Tweets. As before, each suggested positive and negative Tweet was examined and incorrect labels were manually adjusted resulting in 1094 positive and 940 negative Tweets.

The classifier was then retrained over the new training data set using the same features identified earlier. Evaluation of the new classifier using the initial training data as the test set produced an accuracy of 91.13% and an F1 score of 0.921.

4.5 Removing Stop Words

When preparing Tweet ngrams for the classifier, stop words are removed. Our stop word list is similar to those commonly used for traditional Natural Language Processing (NLP) tasks. It has however been extended to include additional Twitter related words and expletives, which are commonly used when experiencing an earthquake event: removing them may reduce the effectiveness of the classifier.

Experiments were run to determine the accuracy of the classifier due to the stop word removal process. Three training and test runs were carried out with various stop words lists: the original list, the original list with expletives and exclamation words removed and an empty list.

The results, shown in Table 3, indicate that the classifier worked slightly better with a modified stop word list and slightly better again with no stop words. Based on this, all future experiments used an empty stop word list.

Table 3: Stop Word Combination Results

Stop word set	F1 score	Accuracy
original	0.9207	91.13%
modified	0.9221	91.19%
empty	0.9223	91.38%

4.6 Feature Selection Revisited

With a now larger collection of annotated Tweet data and evidence that stop word removal should not be used, the feature selection process was revisited. This time, instead of using the 10 fold cross validation process, a simple time-split validation process (Sheridan, 2013) was used: Tweets before a certain time are used to train the classifier and Tweets after are used for testing.

The same cut off time of mid December 2012 was used. In addition to the first set of features evaluated, the presence of a hash tag or mention was now included. This resulted in 144 iterations looking at all feature combinations.

The best performing combination found was unigrams, Tweet length, mention count and hyperlink count with an accuracy of 91.44% and an F1 score of 0.922. Note that in this case, higher accuracy results were achieved without the hash tag related features and unigrams once again outperformed bigrams.

4.7 Tweet Count for Training

Annotating large numbers of Tweets is an onerous process taking considerable time to accomplish. The dependence of the training set size to the accuracy of the resulting classifier was tested. This was done by repeatedly training and testing the classifier using training set sizes of 50, 100, 150 and so on. Initially we naively chose the first 50 Tweets from the training data based on their creation timestamps, and then included the next 50 and so on. The Tweets were selected in chronological order rather than random order to emulate increasing Tweet collection periods. The results are shown in Figure 3.
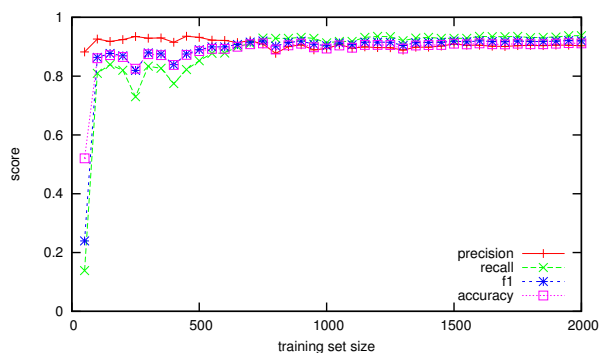

Figure 3: Accuracies: increasing training sizes

Figure 3 shows that almost the same accuracy (89.4% and F1 score of 0.903) can be achieved by annotating only 1000 Tweets which is half the size of our original training set. The variation in the results for smaller training set sizes was concerning. This may be due to a bias resulting from uneven proportions of positive and negative Tweets used. After examining the mix of Tweets used in these test sets, we found this was the case: there were only 71 positive Tweets in the test set of size 500.

To account for the variation in classifier performance over smaller training set sizes, we ran another experiment where we tried to evenly balance the proportion of positive to negative Tweets, where possible. The results are shown in Figure 4 where it can be seen that classifier performance improved significantly even for small data sets.
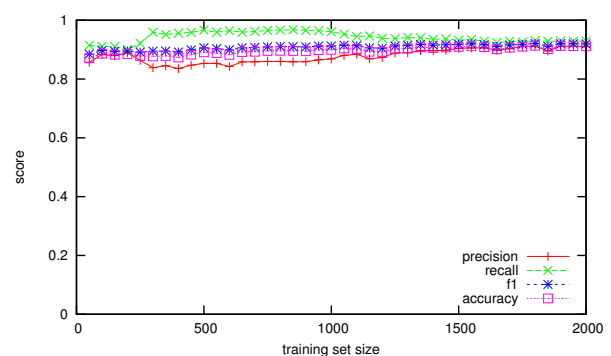

Figure 4: Accuracies: equal Tweet mix

5 Improved Earthquake Alerting

5.1 Summary

The classifier has now been trained on approximately 2000 Tweets from September 2010 to December 2012. It has been configured to use the features: unigrams, Tweet length, hyperlink count and mention count and does not perform stop word removal. Using this classifier, the Tweets contributing to the original post December 2012 earthquake alerts have been reanalysed. For each alert we now generate two additional statistics: the percentage of Tweets classified as positive and the geographic spread (*GeoSpread*) of just the positive Tweets. The *GeoSpread* measure is an indication of how close geographically a collection of Tweets is. This is one of the tests used in the existing heuristic detector and has values ranging from 1 (very close – in the same suburb) through to 15 (far apart – continental scale).

5.2 Further Improvements

When an earthquake alert has subsequent alerts within 30 minutes of the first, the next two alerts

are also used to generate statistics. The aim was to determine if testing follow up alerts is helpful. There have been occasions when the first alert fails the heuristic detector's threshold criteria and is immediately followed by an alert that passes.

We evaluated the alerts using a variety of modifications to the original heuristic algorithm and with variations of classifier configuration. The results are shown in the Table 4.

5.3 Results

As can be seen from Table 4, adding the new rule where the percentage of positively classified Tweets must be at least 50% dramatically increases our F1 score. Also, it has removed all of the original false positive instances. Using the *GeoSpread* of only positively classified Tweets instead of all non-retweets also improved the result, although in our test cases it only removed one false negative instance.

Extending the evaluation to include the alert that immediately follows the original also improved the accuracy. However, in the cases where it is the second alert that passes the test, there is a delay of at least one minute before the notification is sent. This is due to the time taken to identify the second alert generated from the advancing buffer as described in Section 3.1. Evaluating the next two alerts did not improve the accuracy significantly; one false negative instance was removed but a new false positive instance was added.

The final test relaxed the rules for the minimum number of Tweets and the retweet percentage producing the highest F1 score and reducing the number of false negatives to 4, but an extra 2 false positives are introduced.

Overall, the use of a text classifier has greatly improved the accuracy of our earthquake detector, from the original F1 score of 0.667 to 0.881 and original accuracy of 89.87% to 96.85%.

5.4 Deployment

The inclusion of a text classifier has shown to significantly improve the accuracy of our detector. The contents of the notification email generated via the heuristic detector has been extended to include the classification results.

Figure 5 shows an example notification email. The first section contains a summary of the 'earthquake' alert noting the heuristic result that triggered the notification: the *GeoSpread* measures, retweet percentage and a classification summary.

It also contains further information produced from our social media platform not previously discussed: the results of clustering the Tweets contributing to the alert and a summary of the Tweet locations.

```
red 'earthquak' alert generated at: Mon, 6 May 2013 14:20:18 +1000

Statistics:
    Number of tweets (including retweets): 59
    Retweets: 1.69%
    Geographic spread: 0.95

Results of running the "Firsthand earthquake 'felt' reports" classifier:
    Percentage of tweets classified as positive: 91.53%
    Geographic spread of positively classified tweets: 0.95

Location Summary (excluding retweets):
    wellington (-41.28054,174.767136) - 40 tweets
    New Zealand (-43.586578,170.371384) - 11 tweets
    -41.2771551, 174.7774485 (-41.2771551,174.7774485) - 1 tweets
    Aro Valley (-41.296181,174.763046) - 1 tweets
    Petone (-41.221291,174.875076) - 1 tweets
    Upper Hutt (-41.12458,175.069031) - 1 tweets
    Christchurch (-43.531311,172.637299) - 1 tweets
    *unknown location - 2 tweets

Cluster Topics:
    Eqnz - 11 tweets
    Earthquake wellington - 10 tweets
    Earthquakes - 3 tweets
    wind - 3 tweets
    Building - 2 tweets
    Felt - 2 tweets
    Holy - 2 tweets
    Small - 2 tweets
    OMG - 2 tweets
    Other Topics - 33 tweets

Tweets (excluding retweets, +/- labels indicate classification result):
+ 06/05/2013 14:19:36 (wellington, New Zealand) Oooh earthquake.
+ 06/05/2013 14:19:36 (wellington) Earthquake!
+ 06/05/2013 14:19:37 (wellington, New Zealand) Earthquake! #eqnz
+ 06/05/2013 14:19:37 (-41.2771551, 174.7774485) Earthquake?
- 06/05/2013 14:20:12 (wellington New Zealand) eeeek. As if the rain wasn't enough.....an earthquake thrown in.
- 06/05/2013 14:20:15 (wellington, New Zealand) Earthquake?? Everyone round the office thinks so
- 06/05/2013 14:20:16 (Auckland/Christchurch, NZ) Earthquake, Magnitude 3.6, Monday, May 6 2013 at 4:19:11 pm,
30 km south-west of Martinborough http://t.co/XPPrIzdjS2 #eqnz
```

Figure 5: Example notification email

The bottom section of the email summarises the Tweet content each prefixed by a '+' or '-' to indicate the classification result. Note that this example was generated by replaying an historical event and wasn't generated via live Tweet data. Also the list of Tweets has been edited to save space.

The information in the notification email performs three functions: it alerts the recipient of the possibility of an earthquake event, it provides a summary of the reasoning as to why an alert has been generated (the heuristics met and the text classifier results), and it includes a concise summary of the information reported on Twitter.

The reader of the email can quickly assess if the alert is genuine, a *true positive*, and determine the intensity of the earthquake with reference to the number of Tweets reported and by quickly reviewing their content.

6 Further Work

We are developing a classifier to determine an earthquake's intensity analogous to the MMI (Eiby, 1966). Examination of Tweets related to an earthquake reveals that a small percentage contain descriptions of the impact. For example the following (real) Tweet could be classified as level 'VI Strong' in the MMI scale:

Massive earthquake. House covered in glass. Bookshelf on floor. Lights fallen out. Still shaking

This information can be combined with demographic information to help in this determination: a small number of Tweets originating from a sparsely populated region would be given more

Table 4: Results

Modification	TP	FP	TN	FN	F1	accuracy	Heuristic rules used	
Original heuristic	23	11	181	12	0.667	89.87%	numTweets > 3 RT% < 18	geoSpread < 4
Including classification	23	0	192	12	0.793	94.71%	numTweets > 3 RT% < 18	geoSpread < 4 pos% >= 50
Including classification and GeoSpread of positive Tweets	23	0	192	11	0.807	95.13%	numTweets > 3 RT% < 18	posGeoSpread < 4 pos% >= 50
Looking at next alert as well	25	0	192	7	0.877	96.88%	numTweets > 3 RT% < 18 (for either alert 1 or 2)	posGeoSpread < 4 pos% >= 50
Looking at next 2 alerts	25	1	191	6	0.877	96.86%	numTweets > 3 RT% < 18 (for either alert 1 or 2 or 3)	posGeoSpread < 4 pos% >= 50
Looking at the next 2 alerts with relaxed numTweets and RT% rules	26	3	189	4	0.881	96.85%	numTweets > 2 RT% < 30 (for either alert 1 or 2 or 3)	posGeoSpread < 4 pos% >= 50

'weight' compared to the same number of Tweets from a densely populated region. There are other opportunities around data integration also: combining with existing seismic sensor information and utilizing finer grained geo-location of Tweets.

There are also other areas to explore with our Social Media platform. The notification features will be extended to include other emergency management use cases such as fire detection and monitoring, cyclone tracking, flood events and crisis management incidents, for example terrorist attacks and criminal behaviour.

Another area of development is to use classifiers trained to identify impact information. Such classifiers, for example Yin et al. (2012), could be integrated with our system and we intend to experiment with different SVM configurations (error rates and kernel functions) and explore the use of semi-supervised learning using an inductive/transductive SVM to incrementally further train a classifier with user provided input as a 'live' event unfolds.

When an earthquake event is identified, subsequent Tweets could be processed by the impact classifiers to produce a follow up impact analysis email a short time afterwards.

7 Conclusions

Our social media platform provides information captured, filtered and analysed from Twitter using a background language model to characterise the 'normal' activity. Unusual events are identified as alerts when the observed activity varies from that historically recorded. These alerts are then filtered and the contributing Tweets processed to identify evidence of an actual earthquake event.

The initial heuristic based detector has been significantly improved by the introduction of a text classifier. The process of training the classifier has been extensively reported outlining the journey taken to identify different collections of test data, alternative methods of training the classifier, the impact of filtering stop words and the effect of varying the training set size when training the classifier.

The result has been been an incremental improvement to our ability to identify earthquake events as reported on Twitter. Our detector has improved in terms of the *F1 score* from an initial value of 0.667 to 0.881.

Our system generates email notifications of a possible earthquake event, summarises why our system considers it be evidence of firsthand 'felt' reports and includes a concise summary of the information from Twitter. The recipient can quickly assess if the alert is genuine and gain a quick overview of the intensity of the earthquake with reference to the number of Tweets reported and by reviewing their content.

References

Alliance Strategic Research. 2011. Social media in the 2011 victorian floods. Technical report, Office of the Emergency Services Commissioner and Victoria State Emergency Service, GPO Box 4356 Melbourne VIC 3001, June. [Accessed: 6 May 2013].

Martin Anderson. 2012. Integrating social media into traditional management command and control structures: the square peg into the round hole. In Peter Sugg, editor, *Australian and New Zealand Disaster and Emergency Management Conference*, pages 18–34, Brisbane Exhibition and Convention Centre, Brisbane, QLD. AST Management Pty Ltd.

Axel Bruns and Jean E. Burgess. 2012. Local and global responses to disaster: #eqnz and the christchurch earthquake. In Peter Sugg, editor, *Australian and New Zealand Disaster and Emergency Management Conference*, pages 86–103, Brisbane Exhibition and Convention Centre, Brisbane, QLD. AST Management Pty Ltd.

Axel Bruns, Jean Burgess, Kate Crawford, and Frances Shaw. 2012. #qldfloods and @QPSMedia: Crisis Communication on Twitter in the 2011 South East Queensland Floods. Technical report, ARC Centre of Excellence for Creative Industries and Innovation, QUT Z1-515, Musk Ave Kelvin Grove, Qld. 4059 Australia, January.

Mark A. Cameron, Robert Power, Bella Robinson, and Jie Yin. 2012. Emergency situation awareness from twitter for crisis management. In *Proceedings of the 21st international conference companion on World Wide Web*, WWW '12 Companion, pages 695–698, New York, NY, USA. ACM.

Chih-Chung Chang and Chih-Jen Lin. 2011. LIB-SVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Kym Charlton. 2012. Disaster management and social media - a case study. Technical report, Media and Public Affairs Branch, Queensland Police Service, GPO Box 4356 Melbourne VIC 3001. [Accessed: 26 April 2013].

Paul S. Earle, Daniel C. Bowden, and Michelle Guy. 2012. Twitter earthquake detection: earthquake monitoring in a social world. *Annals of GeoPhysics*, 54(6):708–715.

G. A. Eiby. 1966. The modified mercalli scale of earthquake intensity and its use in new zealand. *New Zealand Journal of Geology and Geophysics*, 9(1-2):122–129.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2009. *The elements of statistical learning: data mining, inference and prediction*. Springer, 2 edition.

Thorsten Joachims. 1998. Text categorization with suport vector machines: Learning with many relevant features. In *Proceedings of the 10th European Conference on Machine Learning*, ECML '98, pages 137–142, London, UK, UK. Springer-Verlag.

Bruce R. Lindsay. 2011. Social media and disasters: Current uses, future options, and policy considerations. Technical report, Analyst in American National Government, GPO Box 4356 Melbourne VIC 3001, September. http://www.fas.org/sgp/crs/homesec/R41987.pdf.

Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. 2010. Twitter under crisis: Can we trust what we rt? In *Proceedings of the first workshop on social media analytics*, pages 71–79. ACM.

Qld Budget. 2011. Queensland Government: Budget Strategy and Outlook. page 66, March. [Accessed: 2013-04-9].

RBA. 2011. Reserve Bank of Australia: Statement on Monetary Policy. pages 36–39, February. [Accessed: 2013-04-9].

Bella Robinson, Robert Power, and Mark Cameron. 2013. A sensitive twitter earthquake detector. In *Proceedings of the 22nd international conference on World Wide Web companion*, WWW '13 Companion, pages 999–1002, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 851–860, New York, NY, USA. ACM.

Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2013. Tweet analysis for real-time event detection and earthquake reporting system development. *IEEE Transactions on Knowledge and Data Engineering*, 25(4):919–931.

Robert P. Sheridan. 2013. Time-split cross-validation as a method for estimating the goodness of prospective prediction. *Journal of Chemical Information and Modeling*, 53(4):783–790.

Catherine Stephenson, John Handmer, and Aimee Haywood. 2012. Estimating the net cost of the 2009 black saturday fires to the affected regions. Technical report, RMIT, Bushfire CRC, Victorian DSE, February.

Sudha Verma, Sarah Vieweg, William J Corvey, Leysia Palen, James H Martin, Martha Palmer, Aaron Schram, and Kenneth M Anderson. 2011. Natural language processing to the rescue?: Extracting'situational awareness' tweets during mass emergency. *Proc. ICWSM*.

Jie Yin, Andrew Lampert, Mark Cameron, Bella Robinson, and Robert Power. 2012. Using social media to enhance emergency situation awareness. *IEEE Intelligent Systems*, 27(6):52–59.