AUTOMATIC AND INCREMENTAL KEYWORD EXPANSION FROM SOCIAL MEDIA DATA IN DISASTER

Yi Liu[1], Minsung Hong[2], Rajendra Akerkar[2]

[1] Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

[2] Western Norway Research Institute, P.O.Box 163, NO-6851 Sogndal, Norway

Abstract: *When a disaster occurs, data from social media is one of the essential resources to relevant authorities for the decision making, rescue and replenishment work. Therefore, many researchers have focused on studies to extract more commonly used or disaster-related keywords from the social media data, since people has various ways of expressions and habits to speak about special or day-to-day events. However, existing researches require specific expertise to improve the performance of their approaches. Also, some of them are not able to ensure flexibility of their results according to over time. In particular, most of studies have focused on academic accomplishment than the practical applications. In this paper, we propose an Automatic Disaster Keyword Expansion (ADKE) framework collecting and analysing tweets in real-time to extract localized keywords or general keywords related to disasters. It consists of three modules such as collecting tweets in real-time via Twitter Stream APIs (TSA), identifying localized information using Named Entities Recognition (NER), and topic modelling by Latent Dirichlet allocation (LDA). Based on localized keywords, the proposed framework can be applied to disaster management system. The results of several evaluations by using existing tweet dataset about Tropical Cyclone Oswald in Australia, 2013, showed that the proposed ADKE outperforms the other approach (i.e., pure NER and LDA), without human or expert interventions. We also found that our method combining named entities and keywords for queries to collect tweets via TSA impacts the performance of statistic models of pure LDA.*

Keywords: *Disaster management, Social media, Keyword extraction, Named entity recognition, Topic modelling, Real-time analysis*

1. INTRODUCTION

Disasters such as floods, earthquakes, and terror attacks pose challenges to human society and a threat to people's finances, security, and lives. In addition, disasters are costly in terms of property, loss of life and the consequential instability (Kryvasheyeu et al., 2016). Unfortunately, the frequency and intensity of disasters (both human-made and natural disasters) are increasing significantly due to a range of factors such as extreme climate changes, population growth, and infrastructure aging (Webster, Holland, Curry, & Chang, 2005; Wei, Huang, Lam, & Yuan, 2015). Consequently, disaster damage is expected to be more expensive the foreseeable future. Therefore, disaster resiliency is an important consideration for sustainable development and a priority for urban governments (Wei, Huang, Lam, Sha, & Feng, 2015).

In the era of social media, the use of social media platforms is gaining popularity. For example, Twitter and Facebook have infiltrated people's daily lives. Compared to traditional media channels, social media platforms have clear advantages, such as collecting data in real-time on an unprecedented scale and documenting people's perceptions and reactions to disasters in either virtual or physical environments (Kryvasheyeu et al., 2016). Thus, social media become one of popular data resources. In particular, social media is used widely by people affected by disasters for a variety purposes, including getting updates about the crisis, emergency contacts, personal updates from family and friends, and information about rescue work (Olteanu, Vieweg, & Castillo, 2015). Therefore, many researches using social media data for interpersonal communication (Ludwig, 2017; Whiting & Williams, 2013), citizen sensing (Palen et al., 2010), and official communication in disaster response have been conducted. As one of the important tasks, data acquisition laid the foundation for all progress of leveraging the data. Social media data is usually obtained through Application Programming Interface (API) provided by the service platforms of social media. These APIs give researchers and developers opportunities to obtain messages posted into the platforms.

However, there are two main shortcomings in current methods for collecting social media data. First, current studies are often based on disaster-related expertise, that limits such researches may or may not be available at an emergent disaster period. When a disaster arises, data collection process requires some degree of human intervention to better performance. For example, these methods may require defining specific research terms to use. 'Yin' et al. designed a system leveraging micro-blog data during disasters (Yin et al., 2015), and their data collection needs particular disaster terms as initial condition. On the other hand, gathering social media data tends to be based on static queries which are not adjusted over time. There is a life cycle in disaster response area (Fischer, 1998), which is especially evident in disaster communications. As the changes of disaster phases, different topics are focused and generated. For example, caution and advice messages tend to appear first, and then information about donations or missing people often appears in the days following a disaster (Olteanu et al., 2015). Furthermore, these studies have mainly focused on academic achievement such as improvement of accuracy or precision, rather than applying their approach into real disaster management systems.

To develop a proper approach dealing with these issues, we propose an Automatic Disaster Keyword Expansion (ADKE) framework leveraging tweet data based on a real-world scenario that the tweets are created and collected in real-time. This framework can automatically extract localized and general keywords related to disasters. This framework has three main advantages. One is that our framework using Topic Modelling (TM) techniques to extract keyword. TM is a type of statistical modelling for discovering the abstract "topics" that occur in a collection of documents. In this case, TM is able to help our framework to automatically generate commonly used keywords related to disasters without specific expertise about disasters. Another advantage is that the ADKE is running in real-time and has a self-incremental mechanism. Therefore, the set of keywords is progressively and automatically refined to reflect the disaster evolution. Additionally, the system has strong compatibility in many kinds of disasters. In other words, the proposed framework can be used for various applications. Thus, as shown by results of four systematic experimentations, the proposed framework is well able to be applied to real disaster management systems. The rest of the paper is organized as follows. Section 2 reviews researches related to our framework. The framework is introduced in Section 3, followed by Section 4 discussing the evaluation results. The last section concludes research findings and future work.

2. RELATED WORK

During disasters, a large number of relevant messages are posted to micro-blogging sites, which have led to researches on understanding social media in disasters (Qu, Huang, Zhang, & Zhang, 2011; Starbird, Palen, Hughes, & Vieweg, 2010) and extracting valuable information from it (Imran, Elbassuoni, Castillo, Diaz, & Meier, 2013). The first challenge in using social media data is to retrieve comprehensive collection of disaster-related tweets (Bruns & Liang, 2012). Relevant researches usually used two ways to collect social media data. Keyword-based collection searches and gathers tweets by using a handful of terms and/or hashtags (Bruns, Burgess, Crawford, & Shaw, 2012; Hughes & Palen, 2009). Although the results of the keyword-based collection using hashtags do not have much noise (Vieweg, Hughes, Starbird, & Palen, 2010), they are typically constructed around visible topical hashtags and can omit a large number of disaster-related tweets (Bruns et al., 2012). Whereas keywords are only as responsive as the humans creating them. However, this method might lose relevant tweets because of the latency. As the other way, location-based collection is used to collect only tweets that are either geo-tagged or contain specific places influenced by disasters. Since the location-based method can generally collect small portion of the whole tweets, we basically selected keyword-based collection and use Named Entity Recognition technique to identify location information from the tweets. Detailed methodology is explained in Section 3.

In keyword-based collection, selecting initial keywords is essential. An effective choice of keywords should also be flexible enough to be adapted over time. Analysts can monitor messages posted during a disaster and add/delete keywords to queries as conditions evolve; in fact, experienced analysts may be very skilled at this task. However, in order to reduce the response time and effort required and to improve efficiency, an automatic adjustment and query suggestion for accommodate the running query should be considered. Automatic generation and gradual adjustment of query based on keywords is related to adaptive information filtering, a problem in Information Retrieval, in which we create a dynamic query that is modified as new items of the corpus are discovered (Lanquillon & Renz, 1999), instead of crafting a query for a static collection. Reconstructing the initial query to returns documents from the domain of interest is called vertical selection & aggregation (Arguello, Diaz, & Paiement, 2010) in web search. By focusing on portability and adaptability, 'Arguello' and his colleagues reused past knowledge to predict models for new domains. By following their idea, unlike traditional query generation and expansion on static collections, keywords are automatically and incrementally generated by the proposed framework evolve over time. Furthermore, in contrast to existing approaches utilizing the time dimension of a static micro-blog collection and searching needed data from a historical repository (Metzler, Cai, & Hovy, 2012; Miyanishi, Seki, & Uehara, 2013), we collect data stream in real-time. In order to retrieve more micro-blog data related to given events, 'Wang' et al. expands a user-provided query with new hashtags (Wang, Tokarchuk, Cuadrado, & Poslad, 2013). Whereas the proposed framework automatically retrieves tweets in real-time and extract localized and general keywords which are used to generate a new query for the next cycle of collection.

Once social media data is collected, it is necessary to process the data to convert it into a meaningful information. 'Imran' et al. shown that tweets can contribute to situational awareness, are classified into several types of information (Imran et al., 2013), and the whole process can work automatically. 'Yin' et al. designed a system leveraging social media data during disasters to capture tweets by using a set of specific fixed keywords (Yin et al., 2015). It means their system does not be time-sensitive at disasters. Whereas our framework is progressively and automatically able to refine the keywords in real-time for tweet collecting and can enhance identifying tweets related to disasters. 'Gaglio' et al. presented a system for real-time analysis of tweets in order to track important events from the user's perspective (Gaglio, Re, & Morana, 2016). They also used a set of keywords to generate and refine queries to the new event. The system is similar with our framework; however, the system was designed for general events rather than disaster events and also does not consider localized information. In order to deal with disaster events, our framework deliberately utilizes Named Entity Recognition (NER) and Topic Modelling (TM) and adopts tenable iterative methods on the NER and TM respectively. The proposed approach is able to obtain more localized results to be applied to specific regions or cities.

3. AUTOMATIC DISASTER KEYWORDS EXPANSION

This section explains detailed mechanism of Automatic Disaster Keywords Expansion (ADKE) framework. At the first part we introduce an overview of the framework that consists of three modules which interact dynamically and independently each other to extract a set of localized and general keywords related to a certain disaster. One real-world scenario is also illustrated to explain each module and to describe its function. Then the collaboration among different modules and the dynamic incremental process of about data flow are described.

3.1 ADKE framework

Our framework extracts keywords automatically and adjusts queries, which consists of keywords, over time. Figure 1 shows the structure of the framework in a general perspective. It extracts localized and general keywords, which related to disaster, from tweets data generated in real-time and works with NER and Latent Dirichlet allocation (LDA) repeatedly and continuously to adjust the extracted keywords over time. Three main modules and their functions are given in next sections:

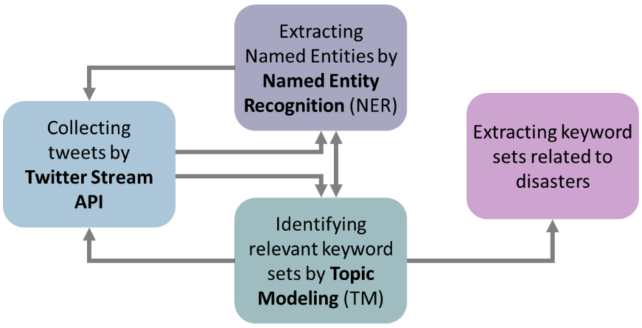

Figure 1 Overview of ADKE framework

3.1.1 Tweets collection module

Tweets collection module is responsible for filter tweets which include specific keyword(s) from whole tweets in real-time, by using Twitter Stream API (TSA). During the execution of this module, tweets including "flood" are continuously being gathered in real-time. Language can be also set as filter conditions. To consider tweets from all over the world, we set the language as English.

3.1.2 Named entity extraction module

Named Entity Recognition (NER) finds Named Entities (NEs) on a text, for example, names of persons, organizations, or places. Table 1 lists different categories of NEs and corresponding explanations.

Table 1 Definition of named entities

Type of Named Entity	**Definition**
NORP	Nationalities, religious or political groups.
FAC	Buildings, airports, highways, bridges, etc.
ORG	Companies, agencies, institutions, etc.
GPE	Countries, cities, states.
LOC	Non-GPE locations, mountain ranges, bodies of water.
EVENT	Named hurricanes, battles, wars, sports events, etc.

In this paper, we use six types of named entities (i.e., NORP, FAC, ORG, GPE, LOC and EVENT). Whereas other NEs are not selected, since they are not sufficiently relevant for the purpose of our framework which take both localized and general keywords into account. Figure 2 shows an example of how the Named Entity Extraction (NEE) module works one tweet related to Queensland flood. By a threshold dynamically adjusted in the proposed framework, it automatically extracts more important NEs based on their occurrence frequency, among all gathered NEs.

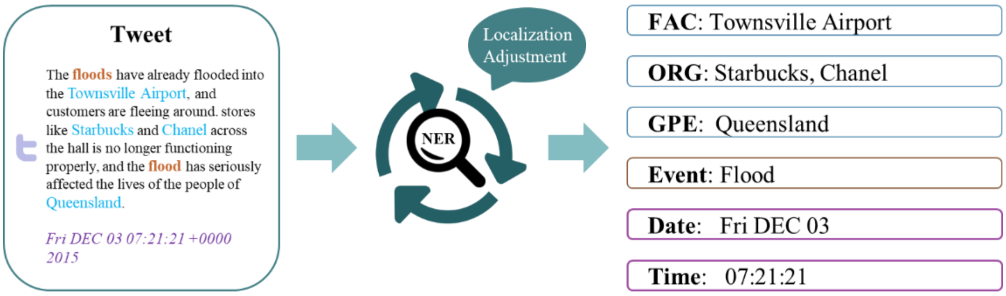

Figure 2 An example of NER process

The NEE module plays an essential role in our framework to locate specific places (e.g., landmarks or stores in the figure above), since the six types of named entities indicate location information. Additionally, the NEE can provide a way to adjust keywords including filtered NEs for Twitter Stream API over time.

3.1.3 Topic modelling module

Latent Dirichlet Allocation (LDA) is often used to create a soft clustering of documents into the re-defined number of topics (BleiDavid, 2003). It assumes that every document reflects a combination of topics, in which the number of relevant topics for a document are a relatively small fraction of all the possible topics. In addition, its basic idea is that every topic can be characterized by a small set of characteristic words that are highly related to that topics, and most of words have the same probability across all topics. Based on these assumptions, LDA can be used as a powerful tool for information extraction in disaster management area using social media data. Therefore, we use LDA as the main engine for keywords extraction in Topic Modelling (TM) module of our framework.

Figure 3 shows an ideal example of how LDA process tweets in the TM module. In this scenario, LDA classifies tweets into several particular topics. However, since LDA can only classify keywords into abstract topics, we should select the important keywords.

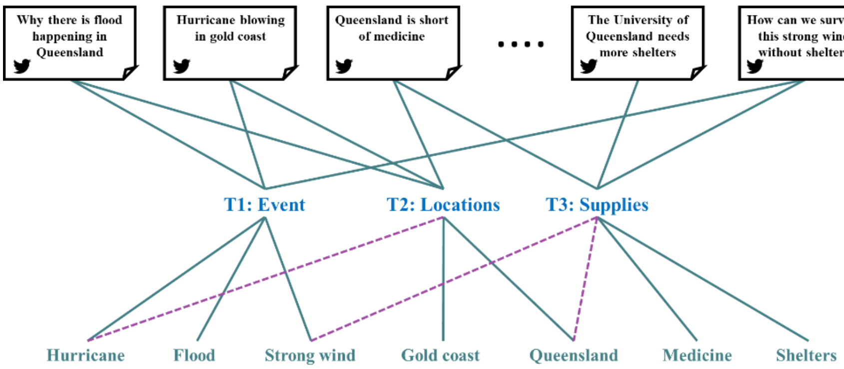

Figure 3 LDA process for tweets related to Queensland flood

Thanks to statistic models (i.e., TF-IDF and Bag of words (BoW)) we can do the task by taking keywords which have highest weights calculated by the models. The weight indicates the importance of a keyword among words in a topic. In addition, even though keyword is not always classified correctly into a corresponding topic like relations which are represented by dashed line, this drawback does not influence our framework. Because there are no strict requirements for matching between topics and keywords for the proposed framework.

3.2 Incremental process of ADKE

In this section, we explain the whole data flow in our framework as shown in Figure 4. The framework initially takes one general keyword as an input. With two different iterations, the framework can generate and adjust the characteristic keywords automatically. The 1st iteration (i.e., NER iteration as dash-dotted line) expands NEs list which emphasizes localized results, since the NEs indicate local information. The 2nd iteration (i.e., topic generation by dashed line) incrementally extracts general annotations related to disasters.

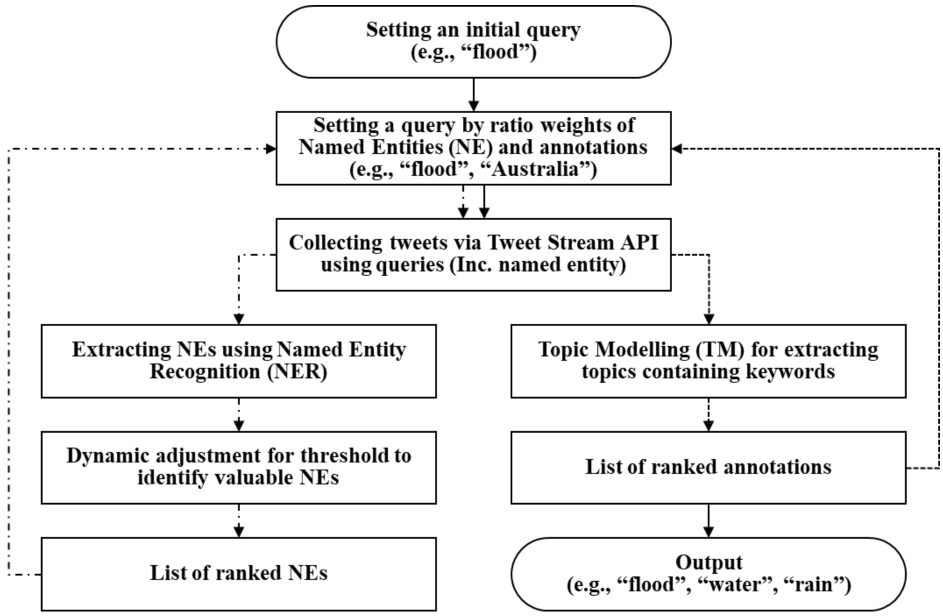

Figure 4 Data flow of AKDE framework

3.2.1 NER iteration

The NER iteration is a combination of Tweets collection module and NER module. The TSA module first takes the initial keyword and filters related tweets posted in real-time. Then the named entity extraction module generates valued NEs based on the occurrence frequency of each NE. Next, all valuable NEs are ranked based on the frequency. The list of ranked NEs is saved as localized information which is updated in every iteration. By using top NEs as a part of the query and continuously executing this iteration over and over again, the list is able to get higher quantity and quality of localized information.

3.2.2 Iteration for Topic generation

As we mentioned in Subsection 3.1.3, topic modelling module digests a large number of tweets as input and generates the most characteristic keywords from the tweets. The output of topic modelling module is a set of topics, and the topics contain characteristic keywords with their importance scores (i.e. weights). We add keywords' importance in different topics individually and generate one annotation list (i.e., set of the keywords with their weights). The total importance of each keyword is used as the standard to order annotations in the list. The top annotation is expected to have more general correlation with a specific disaster. Annotations with high weights are used in the next loop as a query. By this incremental process, the annotation list is expected to be expanded and optimized. Both iterations work simultaneously and independently. In other word, they work without any interference each other, therefore they are able to cooperate together by a simple ratio weights between the numbers of NEs and annotations to generate queries for tweets collection module. The effect of different ratio weights will be evaluated in next section.

4. EVALUATION AND DISCUSSION

The experimentation process and evaluation measurements are explained in this section. Note that we conducted all following experiments in the same computation environment. For the tweets collection module, we used API of Tweepy library[1]. NER and LDA modules were implemented by using NLTK[2] and Gensim[3]. In the first part we describe about dataset used then explain the different static and dynamic methods for thresholds to select valuable NE and its corresponding influence. Performance of eight methods are analysed with various aspects such as the number of iterations, occurrence frequency and so on. In addition, to assess the localized and general annotation extracted by our framework, we compare the lists of NE and keywords using the different weight for control the localization degree in terms of reasonability and logic. Furthermore, with the different weights, we evaluate the effect of two statistic models (Bag of words and TF-IDF) of LDA in our framework by using four measurements (e.g., accuracy, precision, recall and F1-score).

4.1 Dataset

Even though our system operates with data gathered in real-time, it is difficult to conduct evaluations with the data, especially when there is no disaster happening. Therefore, we assessed our framework with an existing labelled dataset which consists of tweets that had been generated during the Tropical Cyclone Oswald which caused serious flood in Queensland of Australia. Therefore, pre-processed the dataset based on its timestamp and synchronized tweets processed for the named entity extraction and topic modelling modules. The dataset has 1,200 tweets while 919 tweets are labelled as related to the corresponding disaster and 281 instances are labelled as unrelated or not applicable tweets. The period is from 17 January 2013 - 5 February 2013.

4.2 Named Entity Filter

From each iteration of NER, a set of NEs is identified. However, not all of them is valuable. Some named entities just occasionally appeared but are not really related to the disaster. To adjust queries, which are used for TSA, with more accurate localized information (i.e., six types of NEs) for the next tweets collection, we need to select relatively more common NEs. As a threshold to determine whether a NE is important, we define the frequency f_{NE_1} of each NE by $\#\ of\ tweets\ containing\ NE_1 / \#\ of\ tweets\ collected.$

By setting a suitable threshold, we can assume that NEs are selected with higher correlation to local disasters. We tested one static and two dynamic methods to set the threshold trying to find the better NEs for queries. The result of this evaluation is summarized in Table 2.

We conducted the evaluation with 8 different conditions and separately recorded the results such as iteration number, average, maximum and minimum numbers of NEs, average frequency and average response time among all iterations. We assume that the higher number of NEs and occurrence frequency would grantee the more localized information (i.e., NEs). To avoid any incorrect assessment of this module, the Topic Modelling module was not used in this evaluation. In addition, NEs ranked within top ten are used as queries from the list of Named Entities for every iteration. If the number of NEs is lesser than ten, then all NEs are used. We set 40 as the number of tweets for one collection circle and processed all 1200 tweets in each condition.

[1] http://docs.tweepy.org/en/latest/#
[2] https://www.nltk.org/
[3] https://radimrehurek.com/gensim/index.html

Table 2 Performance of named entity extraction with different methods

Category	Methods (initial frequency)	Condition(s)	Average No. of the Named Entities	Maximum No. of the Named Entities	Minimum No. of the Named Entities	Average Frequency	Average Response Time (Second)
Static	ST_1 (3%)	-	**4.88**	9	3	0.108	0.730
	ST_2 (7%)	-	3.00	5	1	0.138	0.816
	ST_3 (15%)	-	1.00	2	0	0.200	0.685
Dynamic	DTL_1 (4%)	Min.1, Max.3	2.94	5	0	0.131	0.737
	DTF_1 (7%)	13%	2.88	5	0	0.135	0.854
	DTF_2 (7%)	14%	2.41	5	0	0.148	0.769
	DTF_3 (7%)	20%	1.94	4	0	0.170	0.819
	DTF_4 (7%)	25%	1.50	3	0	**0.191**	0.833

In order to set the threshold, there are three different methods in our evaluation: Static Thresholds (STs), Dynamic Thresholds based on Length of NEs (DTLs) and Dynamic Thresholds based on the Frequency of NEs (DTFs). For the ST method, we set static frequency thresholds (i.e., 3%, 7% and 15%), and only take named entities with higher frequency than the thresholds. We tested each different threshold for whole dataset. To compare with the STs, we used two methods for setting the threshold dynamically. The first DTL's mechanism adjusts frequency thresholds based on the number of NEs. In this method, we can maintain the NEs' number at a proper range. If the number is considered as too many, then the system increases the thresholds 1%. In contrast, if the number of NEs is too small, the thresholds is decreased. In the first dynamic method, a proper number is given by maximum and minimum numbers (i.e., 1 and 3) in Table 2. Whereas the other method (DTF) controls the frequency thresholds according to the average frequency in each iteration. The conditions for the method mean standards to adjust their thresholds. For example, if an initial threshold is 7%, the condition is 13% and the results of average frequency is 14%, then the threshold decreases to 6%. Instead of controlling the number of NEs (i.e., DTL), we choose DTF to more evaluations with different conditions (i.e., 14%, 20% and 25%), since our framework had shown higher frequency performance in our preliminary assessment.

As we increase the static thresholds, the average frequency consequently increased. However, the average number of named entities decreased. Because if we set the thresholds stricter, there are lesser named entities which meet the condition. ST_3 method has shown the highest average frequency among all the static methods as 0.2, however the average number of NEs was 1. Even there was no NE in some iterations. Whereas ST_1 method had the highest average number of NEs, but its frequency was lowest among static methods. Therefore, there is need to use a dynamic method to adjust thresholds with expecting that a dynamic method, especially average frequency, can reach the frequency level of ST_3 with more NEs than static methods. As shown by the result of DTF_4, its average frequency was 0.191, a litter lower than 0.2. Furthermore, the average number of named entities was 1.5. In other words, this result outperforms than ST_3 as 50% higher, which means while this dynamic method can not only achieve the similar quality of named entities like static methods but also higher quantities of the named entities. Due to this self-regulation, dynamic methods based on average frequency are more flexible for processing real-time data in the same mechanism in real systems (i.e., using Twitter Stream API). Therefore, we adopted this method into our framework and used for next evaluations.

4.3 Localization and general results

As aforementioned, the named entity extraction module is in charge of identifying localized information. And the information feeds to queries for collecting tweets which are used in Topic Modelling (TM) module focusing on general keyword extraction. In addition, the keywords from the list of ranked annotations are also leveraged in the collection module with the NEs from the list of ranked NEs as shown in Figure 5. Therefore, we need a proper weight to combine both results into generating queries. In order to assess the weight, we set different proportion between NEs and the keywords, called annotations, for the weight. The weight w_{NEA} is defined by $\#\ of\ Named\ Entitie / (\#\ of\ Named\ Entitie\ + \#\ of\ Annotation)$.

Table 3 and Table 4 show the lists of NEs and top 10 annotation which are obtained by using different w_{NEA} (0, 0.5 and 1), respectively. Using the weight as 0 indicate that we set our framework as leveraging only the pure LDA, while for 1, it means we use only the pure NER. Note that the annotation list is generated by whole dataset, since TM uses accumulated tweets. If the w_{NEA} is 0.2, three NEs and eight annotated keywords are contained into a query. It may mean that the tweets filtered by the query would have more localized information because of NEs which belong to lexicon represents places. The score in Table 4 means importance of keywords among all topics. Results of LDA are represented by several topics including a set of keywords.

Table 3 List of named entities per weight

$w_{NEA} = 0$ (pure LDA)		$w_{NEA} = 0.5$ (ADKE)		$w_{NEA} = 1$ (pure NER)	
Named Entity	**Average Frequency**	**Named Entity**	**Average Frequency**	**Named Entity**	**Average Frequency**
Australia	0.124	Australia	0.113	Australia	0.152
Queensland	0.076	Queensland	0.085	Queensland	0.103
QLD	0.016	QLD	0.036	QLD	0.028
Gold Coast	0.01	Gold Coast	0.011	the Coral Sea	0.013
the Coral Sea	0.008	the Coral Sea	0.010	the Space Station	0.013
the Space Station	0.008	the Space Station	0.010		
Australian	0.005	Australian	0.005		
BBC News	0.004				
Total average frequency: 0.0313		Total average frequency: 0.0383		Total average frequency: 0.0616	
# of tweets: 820		# of tweets: **842**		# of tweets: 667	

When we compare the results of pure LDA ($w_{NEA} = 0$), pure NER ($w_{NEA} = 1$) and combination of both techniques equally, the result of the pure LDA contained un-localised information such as BBC News. Regarding the maximum number of NEs as shown in 1st evaluation is 9, one irrelevant NE has quite big impact as 11%. By comparing with the combined method (i.e., ADKE), the pure NER missed two localized information. On the other hand, all NEs from the combination method are localized information. In terms of total average frequency and the numbers of tweets used, the combination method still outperforms the others. Even though the frequency of the combination method is much lower than that of pure NER, the number of tweets, which can be used, for the ADKE are larger than NER's one. Regarding the whole results, the combination method with a proper weight is the best. As shown in Table 4, we are able to say that the combination method (i.e., ADKE) outperforms the others (i.e., pure LDA and NER) in terms of the number of relevant keywords for annotations.

Table 4 List of annotations per weight

Weight (w_{NEA})	0 (pure LDA)	0.5 (ADKE)	1 (pure NER)
Total # of keywords	46	47	30
# of irrelevant keywords	19	19	10
# of relevant keywords	27	**28**	20
Ratio of relevant keywords	0.5870	0.5957	0.6667

In other words, our framework collects more relevant keywords than the others. In this evaluation, we used Bag of Words (ToW) technique as statistic model for LDA. This result accords with the ultimate goal of the proposed ADKE framework. The original and valid results of each weight are given in Table 5. Note that we had manually validated whether a keyword is directly related the disaster Tropical Cyclone Oswald or not. Irrelevant keywords are indicated in bold.

Table 5 List of detailed annotation results

w_{NEA}=0 (pure LDA)				**w_{NEA}=0.5 (ADKE)**				**w_{NEA}=1 (pure NER)**			
Original results		**Valid results**		**Original results**		**Valid results**		**Original results**		**Valid results**	
Keyword	Score	Keyword	Score	Keyword	Score	Keyword	Score	Keyword	Score	Keyword	Score
bigwet	1.85	bigwet	1.85	bigwet	1.252	bigwet	1.252	**http**	1.554	bigwet	1.283
qldflood	0.906	qldflood	0.906	qldflood	0.894	qldflood	0.894	bigwet	1.283	qldflood	1.213
brisban	0.345	brisban	0.345	brisban	0.286	brisban	0.286	qldflood	1.213	brisban	0.305
newsbrisban	0.161	coast	0.148	affect	0.226	affect	0.226	brisban	0.305	safe	0.208
coast	0.148	water	0.124	**love**	0.202	safe	0.138	**peopl**	0.245	affect	0.196
love	0.144	river	0.116	**news**	0.172	coast	0.126	safe	0.208	help	0.137
water	0.124	affect	0.097	**peopl**	0.147	rise	0.114	affect	0.196	tornado	0.136
laugh	0.12	dead	0.095	safe	0.138	bundaberg	0.1	help	0.137	rain	0.132
river	0.116	tornado	0.091	coast	0.126	water	0.1	tornado	0.136	rise	0.131
affect	0.097	bundaberg	0.088	rise	0.114	dead	0.098	rain	0.132	coast	0.131
abcnew	0.097	safe	0.079	**stay**	0.111	storm	0.096	**news**	0.132	emerg	0.124
dead	0.095	power	0.075	**today**	0.107	heavi	0.095	rise	0.131	bundaberg	0.121
today	0.094	rise	0.073	bundaberg	0.1	rain	0.087	coast	0.131	dead	0.105
peopl	0.092	help	0.071	water	0.1	tornado	0.085	**love**	0.13	water	0.099
tornado	0.091	live	0.069	dead	0.098	help	0.079	emerg	0.124	australian	0.096
bundaberg	0.088	warn	0.058	storm	0.096	weather	0.074	**newsbrisban**	0.124	alert	0.095
gold	0.085	need	0.057	heavi	0.095	cyclon	0.07	bundaberg	0.121	evacu	0.092
news	0.082	heavi	0.057	**gold**	0.091	good	0.067	**stay**	0.117	live	0.089
safe	0.079	alert	0.056	rain	0.087	evacu	0.064	dead	0.105	warn	0.082
power	0.075	good	0.055	tornado	0.085	alert	0.06	water	0.099	river	0.072
rise	0.073	disast	0.044	**abcnew**	0.084	emerg	0.057	australian	0.096		
photo	0.073	evacu	0.043	**state**	0.082	river	0.048	alert	0.095		
help	0.071	rain	0.042	help	0.079	live	0.043	evacu	0.092		
live	0.069	storm	0.039	weather	0.074	home	0.041	live	0.089		
state	0.066	cyclon	0.037	**laugh**	0.071	week	0.033	warn	0.082		
like	0.063	emerg	0.034	**know**	0.07	warn	0.029	**state**	0.079		
clean	0.059	weather	0.033	cyclon	0.07	power	0.027	river	0.072		
warn	0.058			good	0.067	need	0.019	**updat**	0.063		
updat	0.057			evacu	0.064			**today**	0.06		
need	0.057			**photo**	0.061			**abcnew**	0.059		
heavi	0.057			alert	0.06						
go	0.056			**go**	0.06						
qpsmedia	0.055			**clean**	0.059						
alert	0.055			emerg	0.057						
time	0.054			**newsbrisban**	0.055						
know	0.054			**like**	0.053						
good	0.044			river	0.048						
disast	0.043			**updat**	0.044						
evacu	0.042			live	0.043						
thought	0.04			home	0.041						
stay	0.039			**time**	0.035						

rain	0.039	**auspol**	0.035
storm	0.037	week	0.033
cyclon	0.034	warn	0.029
emerg	0.033	**qpsmedia**	0.028
weather	0.031	power	0.027
		need	0.019

As shown in Table 3 and Table 4, the Named Entities Extraction module and Topic Modelling module show quite different output. The list of NEs shows a fairly localized information such as "Australia" and "Queensland" while the annotation list focused more on general keywords related to flood. In terms of extracting localized and general annotations, our ADKE framework combining LDA and NER techniques archived best performance by comparing with pure LDA and NER. Notably we should emphasize that the framework is able to work in real system, since all evaluations using existing dataset have been conducted by the same mechanism including use of TSA which collects tweets in real-time.

In order to observe the influence of the proposed ADKE framework to LDA in detail, we conducted another evaluation using two different statistic models for LDA. With different weights, the performance of BoW and TF-IDF differs. In our evaluation, we used the combined query to filter whole datasets. Data cluster containing the query will be recognized as disaster-related cluster. The other cluster will be recognized as un-related cluster. The number of tweets which are labelled as "related" in disaster related cluster is the parameter True Positives (TP), while the number of tweets labelled in opposite way is False Positives (FP). In the un-related cluster, the quantity of tweets labelled as "related" is False Negatives (FN), and in the opposite case we can get True negatives (TN). Based on these four parameters, we used four measurements (e.g., Accuracy, Precision, Recall and F1-score) to assess them. The results of statistic modes according to query generation are listed in Table 6.

Table 6 Performance of statistic models according to query generation

w_{NEA}	**Statistic Model**	**Accuracy**	**Precision**	**Recall**	**F1-score**	w_{NEA}	**Statistic Model**	**Accuracy**	**Precision**	**Recall**	**F1-score**
0	BoW	0.7018	0.7529	0.9083	0.8234	0.6	BoW	**0.7268**	0.7590	**0.9421**	**0.8407**
	TF-IDF	0.7034	0.7534	0.9105	0.8245		TF-IDF	0.7185	0.7569	0.9312	0.8350
0.1	BoW	0.7026	0.7532	0.9094	0.8239	0.7	BoW	0.7185	0.7569	0.9312	0.8350
	TF-IDF	0.7034	0.7534	0.9105	0.8245		TF-IDF	0.7185	0.7569	0.9312	0.8350
0.2	BoW	0.7068	0.7543	0.9148	0.8268	0.8	BoW	0.7185	0.7573	0.9301	0.8349
	TF-IDF	0.7018	0.7530	0.9083	0.8234		TF-IDF	0.7185	0.7569	0.9312	0.8350
0.3	BoW	0.7201	0.7573	0.9334	0.8362	0.9	BoW	0.7185	0.7569	0.9312	0.8350
	TF-IDF	0.7193	0.7571	0.9323	0.8356		TF-IDF	0.7185	0.7573	0.9301	0.8349
0.4	BoW	0.725	0.7586	0.9400	0.8396	1	BoW	0.6642	**0.8671**	0.6627	0.7512
	TF-IDF	0.7201	0.7573	0.9334	0.8362		TF-IDF	0.6642	**0.8671**	0.6627	0.7512
0.5	BoW	0.7201	0.7573	0.9334	0.8362						
	TF-IDF	0.7185	0.7569	0.9312	0.8350						

BoW has generally better performance than the TF-IDF in our experimentation, expect for with the weight 0 (i.e., pure LDA). Furthermore, the performance of BoW with the weight 0.6 is the best in terms of accuracy, recall and F1-score. In other words, our combination method outperforms normal LDA using TF-IDF in terms of generating annotation. However, when we set the weight as too high such as 0.9 or 1, the difference between BoW and TF-IDF almost disappears. The reason of these results might be that the two statistic models are parts of Topic Modelling module, that is, if influence of Topic Modelling module is decreased, the difference between Bag of words and TF-IDF will also decrease.

5. CONCLUSION

Recent days, social media data has become more crucial in disaster decision making. Therefore, many researchers have tried to extract more important information such as relevant keywords, people's needs and so on from the data. However, most of studies requires expertise to improve accuracy, and some of them are not feasible to adapt their result over time. Furthermore, these researches seem to focus more on academic accomplishment, rather than consideration of applicability to real disaster management systems. In this paper, we propose an Automatic Disaster Keyword Expansion (ADKE) using Twitter Stream APIs (TSAs), Named Entities Recognition (NER) and Latent Dirichlet allocation (LDA) in order to extract localized keywords or general keywords related to disasters. By localized keywords from tweets, our framework is more able to be applied to real disaster management systems. The ADKE framework generates automatically and incrementally these keywords over time. In other words, the framework can gradually work without human interventions and expertise. To show the effectiveness of our approach, four experiments were conducted with existing tweet dataset about Tropical Cyclone Oswald in Australia, 2013. Note that the mechanism of the approach is exactly same with real-time analysis using Twitter Stream APIs. That is, the mechanism can be applied to real disaster management systems. Results from our evaluation showed that the proposed framework outperforms pure NER and LDA for both localized and general keyword extractions. Also, we found that our method for generating queries to collect tweets via TSA impacts the performance of statistic models of pure LDA. The results from the four systematic evaluations showed that our framework is applicable to disaster management system in social media generation. In this paper, our main contribution is design, implementation, and demonstration of the proposed ADKE framework which performs automatically and incrementally.

As future work, we will implement an additional module to automatically obtain authoritative location data from systems of certain regions or cities, and develop a dynamic way to set the ratio between named entities and general keyword for queries used for real-time collecting tweets.

Acknowledgement

The work is funded from the Research Council of Norway (RCN) and the Norwegian Centre for International Cooperation in Education (SiU) grant through INTPART program.

Reference

Arguello, J., Diaz, F., & Paiement, J.-F. (2010). *Vertical selection in the presence of unlabeled verticals.* Paper presented at the Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval.

Bruns, A., Burgess, J., Crawford, K., & Shaw, F. (2012). Crisis communication on Twitter in the 2011 South East Queensland floods.

Bruns, A., & Liang, Y. E. (2012). Tools and methods for capturing Twitter data during natural disasters. *First Monday,* **17**(4), 1-8.

Fischer, H. W. (1998). *Response to disaster: Fact versus fiction & its perpetuation: The sociology of disaster*: University press of America.

Gaglio, S., Re, G. L., & Morana, M. (2016). A framework for real-time Twitter data analysis. *Computer Communications,* **73**, 236-242.

Hughes, A. L., & Palen, L. (2009). Twitter adoption and use in mass convergence and emergency events. *International journal of emergency management,* **6**(3-4), 248-260.

Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., & Meier, P. (2013). *Practical extraction of disaster-relevant information from social media.* Paper presented at the Proceedings of the 22nd International Conference on World Wide Web.

Kryvasheyeu, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., & Cebrian, M. (2016). Rapid assessment of disaster damage using social media activity. *Science advances,* **2**(3), e1500779.

Lanquillon, C., & Renz, I. (1999). *Adaptive information filtering: Detecting changes in text streams.* Paper presented at the Proceedings of the eighth international conference on Information and knowledge management.

Ludwig, T. (2017). Dynamic Quality Assessment of Citizen-Generated Content (TOCHI). In *Researching Complex Information Infrastructures* (pp. 127-163): Springer.

Metzler, D., Cai, C., & Hovy, E. (2012). *Structured event retrieval over microblog archives.* Paper presented at the Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.

Miyanishi, T., Seki, K., & Uehara, K. (2013). *Improving pseudo-relevance feedback via tweet selection.* Paper presented at the Proceedings of the 22nd ACM international conference on Information & Knowledge Management.

Olteanu, A., Vieweg, S., & Castillo, C. (2015). *What to expect when the unexpected happens: Social media communications across crises.* Paper presented at the Proceedings of the 18th ACM conference on computer supported cooperative work & social computing.

Palen, L., Anderson, K. M., Mark, G., Martin, J., Sicker, D., Palmer, M., & Grunwald, D. (2010). *A vision for technology-mediated support for public participation & assistance in mass emergencies & disasters.* Paper presented at the Proceedings of the 2010 ACM-BCS visions of computer science conference.

Qu, Y., Huang, C., Zhang, P., & Zhang, J. (2011). *Microblogging after a major disaster in China: a case study of the 2010 Yushu earthquake.* Paper presented at the Proceedings of the ACM 2011 conference on Computer supported cooperative work.

Starbird, K., Palen, L., Hughes, A. L., & Vieweg, S. (2010). *Chatter on the red: what hazards threat reveals about the social life of microblogged information.* Paper presented at the Proceedings of the 2010 ACM conference on Computer supported cooperative work.

Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010). *Microblogging during two natural hazards events: what twitter may contribute to situational awareness.* Paper presented at the Proceedings of the SIGCHI conference on human factors in computing systems.

Wang, X., Tokarchuk, L., Cuadrado, F., & Poslad, S. (2013). *Exploiting hashtags for adaptive microblog crawling.* Paper presented at the Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining.

Webster, P. J., Holland, G. J., Curry, J. A., & Chang, H.-R. (2005). Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science,* **309**(5742), 1844-1846.

Wei, Y., Huang, C., Lam, P., Sha, Y., & Feng, Y. (2015). Using urban-carrying capacity as a benchmark for sustainable urban development: an empirical study of Beijing. *Sustainability, 7*(3), 3244-3268.

Wei, Y., Huang, C., Lam, P. T., & Yuan, Z. (2015). Sustainable urban development: A review on urban carrying capacity assessment. *Habitat International*, **46**, 64-71.

Whiting, A., & Williams, D. (2013). Why people use social media: a uses and gratifications approach. *Qualitative Market Research: An International Journal*, **16**(4), 362-369.

Yin, J., Karimi, S., Lampert, A., Cameron, M., Robinson, B., & Power, R. (2015). *Using social media to enhance emergency situation awareness.* Paper presented at the Twenty-fourth international joint conference on artificial intelligence.