# Classifying Micro-text Document Datasets: Application to Query Expansion of Crisis-Related Tweets

Mehrdad Farokhnejad, Raj Ratn Pranesh, Javier A Espinosa-Oviedo

# Classifying Micro-text Document Datasets: Application to Query Expansion of Crisis-Related Tweets*

Mehrdad Farokhnejad[1], Raj Ratn Pranesh[2], and Javier A. Espinosa-Oviedo[3]

[1] Univ. Grenoble Alpes, Grenoble INP, CNRS, LIG, France
Mehrdad.Farokhnejad@univ-grenoble-alpes.fr
[2] Birla Institute of Technology, Mesra, India
raj.ratn18@gmail.com
[3] University of Lyon, LIG-LAFMIA, France
javier.espinosa-oviedo@univ-lyon2.fr

**Abstract.** Twitter is an active communication channel for spreading information during crises (e.g., earthquake). To exploit this information, civilians require to explore the tweets produced along a crisis period. For instance, for getting information about crisis' related events (e.g. landslide, building collapse), and their associated relief actions (e.g., gathering of food supply, search for victims). However, such Twitter usage demand significant effort and answers must be accurate to support the coordination of actions in response to crisis events (e.g., avoiding a massive concentration of efforts in only one place). This requirement calls for efficient information classification so that people can perform agile and useful relief actions. This paper introduces an approach based on classification and query expansion techniques in the context of micro-texts (i.e., tweets) search. In our approach, a user's query is rewritten using a classified vocabulary derived from top-k results, to reflect her search intent better. For classification purpose, we study and compare different models to find the one that can best provide answers to a user query. Our experimental results show that the use of Multi-Task Deep Neural Network (MT-DNN) models further improves micro-text classification. Also, the experimental results demonstrate that our query expansion method is effective and reduces noise in the expanded query terms when looking for crisis tweets on Twitter datasets.

**Keywords:** Crisis Computing · Tweets Classification · Query expansion · Microblog Retrieval

## 1 Introduction

With the rapid development of social networks and the Internet, social media are being used more and more for communicating, tracking, and extracting

* This work was partially funded by the Iranian Ministry of Science, Research and Technology through the fellowship of Mehrdad Farokhnejad.

information about currently occurring or recently passed crises. However, important information is hidden within a large volume of irrelevant and noisy content[16]. Twitter is one of the popular microblog service providers which people at the scene of a disaster post information about the disaster on it. This citizen-generated data provides information about the need and availability of resources at the affected locations which humanitarian organizations can use this information to provide relief. For instance, for getting information about crisis' related events (e.g. landslide, building collapse), and their associated relief actions (e.g., gathering of food supply, search for victims). However, such Twitter usage demand significant effort and answers must be accurate to support the coordination of actions in response to crisis events (e.g., avoiding a massive concentration of efforts in only one place). This requirement calls for efficient information classification so that people can perform agile and useful relief actions.

Despite the potential benefits, it is increasingly difficult to accurately and thoroughly obtain useful information from massive microblog datasets using traditional information retrieval models. The size of the microblog texts that contain few semantic information increases the difficulty of analysing their content. Also, since microblog retrieval only uses the search keywords provided by the users, there is a considerable risk that query terms fail to match any word observed in relevant tweets. The existing research shows that searchers supply two or three query terms on average [18], which is a short number and these terms can only express a small part of the user's information needs. To overcome mismatch problem (i.e., how to retrieve concise documents, which might be conceptually relevant, but do not explicitly contain some or all of the query terms) query expansion techniques [21] provide alternatives.

Query expansion (QE) techniques refer to the process of reformulating queries with additional terms that better define the information needs of the user [1]. Query expansion approaches rewrite the original query by adding other relevant keywords or suggesting additional appropriate keywords. Classical QE techniques have achieved good results in traditional text retrieval tasks. However, directly applying these methods to microblog information retrieval cannot achieve the desired performance, given the characteristics of the posts [21, 12]. To overcome the limitations of existing methods, we propose a classification based method to extract relevant keywords for query expansion. In this method, the aim is to find expanded query terms from top documents. To do this, we firstly preprocess an initial query and classify it; then we extract frequent terms from top results. The idea is that we only consider top results that are in the same class as a query to extract the expanded terms. For classification purpose, we study and compare different models namely, Support Vector Machines (SVM), Naive Bayes (NB) and Random Forest (RF); the convolutional neural model (CNN) and the Multi-Task Deep Neural Network (MT-DNN) to find the best model for classifying crisis micro-texts. We conducted experiments to assess the models on a crisis tweet dataset [7, 14].

The main contributions of this paper include: (1) the leverage of a novel approach to extract relevant terms to expand the original query, which can better reflect users' search intent, and (2) the implementation and comparison of several classification techniques. The objective was to study and compare different models for addressing the classification of crisis-related tweets. We show that we obtain competitive results with other works addressing crisis tweets analysis [7, 14], and we achieve to obtain better results applying MT-DNN for tweets classification in a novel and original manner. Besides, the experimental results demonstrate that our query expansion method is effective and reduces noise in the expanded query terms, which improves the accuracy of microblog retrieval.

The remainder of the paper is organized as follows. Section 2 discusses related work and compares our work with approaches addressing crisis micro-text classification and query expansion. Section 3 describes the general micro-texts classification study expressed as a classification problem addressed using supervised and deep learning models. Section 4 describes the experimental settings, datasets and discusses the obtained results. Section 5 concludes the paper and discusses future work.

## 2 Related Work

Millions of people use social media platforms, and the amount of data they produce is enormous. Researchers are continuously working on developing systems which can efficiently process the human-generated data during events like disasters to use them for building solutions that can save millions of lives.

Studies have shown how crisis data can be beneficial and crucial in analysing and collecting insight during and after a disaster. In the paper [19], the authors proposed a match discovering system for mapping the disaster aid messages and victims problem reports. Authors in the [2], analysed the social media data generated during the occurrence of a disaster. Paper [17] presented a classification system for identifying the type of disaster tweet. Research in the field of query processing can be classified, based on the source of expansion terms, into three groups: query expansion based on relevance feedback, query expansion based on local analysis, and query expansion based on global analysis [22]. Query expansion based on relevance feedback utilises feedback from the initial retrieval to enrich the original query. Query expansion based on local analysis is also known as pseudo-relevance feedback method. Specifically, the retrieval system assumes that the first k documents returned are relevant documents and query expansion words extract from the top k retrieved documents. Query expansion based on global analysis aims to mine the relevance difference among words, and treats the most relevant words as complements to the query.

The traditional text retrieval field applies the query mentioned above expansion methods. However, it is not easy to achieve the desired performance by directly using these methods in microblog retrieval [10, 22]. The reason is that there is a large number of network vocabularies in microblogs and the junk text, without any useful information. Because of these factors, if top-ranked microblogs, returned by the initial search, are not relevant, microblog query expansion through pseudo-relevance feedback will be of little use.

Our work integrates a tweet classification process to retrieve documents that are in the query's class to extract relevant keywords for query expansion. For classifying crisis-related data, various machine learning algorithms and their performance have been proposed [7, 8, 5]. In [6], authors have shown DNN outperforms the traditional models in most of the tasks. The results of applying CNN for analysing crisis data [14, 13] have surpassed the traditional machine learning models by a significant margin. The authors proposed the semantically-enhanced duel-CNN with two layers in [4].

Our work applied and compared the techniques previously used to classify micro-texts, and particularly crisis tweets related to disasters. We reproduced existing experiments like [7, 14]. Seeking for better performance with the datasets we used, we applied MT-DNN. The application of MT-DNN in this context is novel and original and has led to promising results. Moreover, based on our preliminary experiments, we observed that query expansion based on the classification method obtains better candidate expansion words, which are semantically close to the user query.

## 3 Query Expansion Based On Classification Results

Our approach is calibrated to explore disaster management datasets (e.g. earthquake, flooding, fire) produced by social media. We use prepared tweet disaster datasets ready to be explored. We focused on expanding queries looking for micro-texts (i.e., tweets) related of two classes: events which represent situations produced during the disaster life cycle (e.g., someone looks for shelter, a building has been damaged); and actions performed in response to events (a hotel is providing shelter for victims, people is approaching a damaged building to search victims).

Figure 1 illustrates the proposed framework to find the relevant terms to expand a query. It consists of two phases. The first phase pre-processes an initial query to rewrite it by extending it with relevant terms. The second is devoted to classifying the query to determine its type (i.e., event, action) inspired in [15, 20].

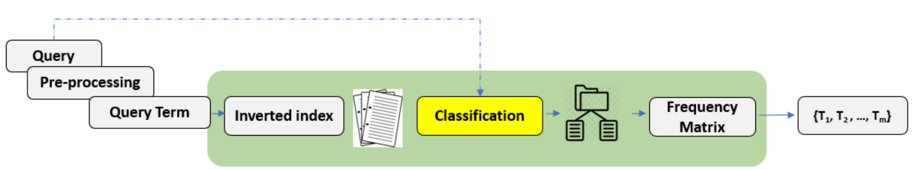


**Fig. 1.** Classification Based Query Expansion

### 3.1 Phase A

Phase (**A**) consists of the following steps:

(i) The original query is preprocessed and cleaned removing stop words and symbols. (ii) Then the query is classified into two classes (event, action). The classified query is used to obtain a set of relevant tweets from a large unlabelled tweet corpus using an inverted indexed matrix consisting of terms extracted from the tweet corpus. (iii) Once we have a set of relevant tweets, we classify them using our classification language models and select the tweets that belong to the user query class. (iv) We obtain the **m** top frequent keywords out of the classified relevant tweets (in step (iii)). We have elaborated our detailed approach for phase A in algorithm 1. The following paragraphs give details of the most relevant steps of the algorithm.

**Algorithm 1:** Finding Relevant Terms

```
Data: (Crisis Related Tweets, User Query)
Result: Expanded Query Terms
begin
    Cleaned_Data ⟵ Data_Cleaning(CrisisRelatedTweets)
    Indexed_Data ⟵ Inverted_Index(Cleaned_Data)
    Query_Keywords ⟵ PreProcessing(UserQuery)
    for Keyword ∈ Query_Keywords do
        Initial_Result ⟵ Finding_tweets_contain_keyword(Keyword)
    Query_Label ⟵ MTDNN_Classifier(UserQuery)
    for Tweet ∈ Initial_Result do
        if MTDNN_Classifier(Tweet) == Query_Label then
            Final_Result ⟵ Tweet
    ExpandedQueryTerms ⟵ Term_Frequency(Final_Result)
```

**Classifying Micro-Texts.** The data science workflow implementing the classification phase applies different machine learning and deep learning models [7, 14]. The workflow splits into three groups of activities: (1) data preparation; (2) classification and (3) assessment. The activities of group 2 and 3 are specialised into the following activities: (2.1) Creation of a baseline applying supervised learning models (i.e., Support Vector Machines (SVM), Random Forest(RF) and Naive Bayes (NB) as classic classifiers) and (3.1) their assessment. (2.2) Classification with no prior knowledge and (3.2) assessment. (2.3) Classification based on MT-DNN that looks for a better classification score and (3.3) assessment. Assessment activities enable the comparison of the performance of the models according to their accuracy, to choose the one that provides the best results for rewriting the queries.

*Classification baseline.* As said before, for the classification step, our objective was to identify 2 classes: events (situations coming up in disaster) and actions (reactions performed in response to events).

We first implemented a supervised learning classification that maps an input (tweet) to an output (label) based on labelled crisis data available on crisis NLP website [7]. For example, the tweet "#BREAKING New Injury Numbers 172 injured, 7 fractures, 1 critical #napaquake" is related to the concepts of death and accident, so it is mapped to the class *event*. In contrast, the tweet "Full statement by Napa Valley Vintnerson new #earthquake relief find, with a link for making donations." concerns Non-Governmental Organisations (NGO) efforts and donations, so it is mapped to the class *action*. We used three supervised learning algorithms, namely, Support Vector Machines (SVM), Random Forest(RF) and Naive Bayes (NB) as classic classifiers.

*Convolutional Neural Networks.* The activities of the data science workflow specialised on Deep neural networks (DNNs)[9] were designed as follows. We used a Convolutional Neural Network (CNN) which a deep learning network consisting of an input layer, multiple convolution layers and an output layer. For applying CNN in NLP tasks, like tweets classification, we used previously computed token sequences as input to the CNN. Then, CNN filters preform as n-grams over continuous representations. These n-grams filters are combined by subsequent network layers, namely the dense layers. CNN can learn the features and distinguish them automatically, and therefore, it does not require hand-engineered features. This saves human effort and time and eliminates the need for prior knowledge. A distributed word representation and generalisation feature effectively utilise the already used labelled data from the other event. This increases the efficiency of the classification process on new data. It removes the need to use manually craft features as it learns automatically latent features as distributed dense vectors, which generalise well and have shown to benefit various NLP tasks[14].

*Looking for better classification results.* We used Multi-Task Deep Neural Networks (MT-DNN) [11] to classify the tweets looking for better classification results. MT-DNN is based on knowledge distillation which is a process of transferring the knowledge from a set of the larger, complicated model(s) to a lighter compact, easier to deploy single model, without significant loss in performance. In MT-DNN Lexicon Encoder $l_1$ and Transformer Encoder $l_2$ are the shared layer. The input sentence: $X = \{x_1, x_2, \cdots, x_m\}$ is a sequence of tokens of length m. Then the lexicon encoder maps X into a sequence of input embedding vectors, one for each token, constructed by summing the corresponding word, segment, and positional embeddings. In layer $l_2$ a multilayer bidirectional Transformer encoder is used to map the input representation vectors $l_1$ into a sequence of contextual embedding vectors. MT-DNN learns the representation using multi-task objectives, in addition to pre-training. The MT_DNN model is shown in figure 2.

We fine-tuned the MT-DNN codebase to perform the specific task of single sentence classification. The input X (a sentence) is first represented as a sequence

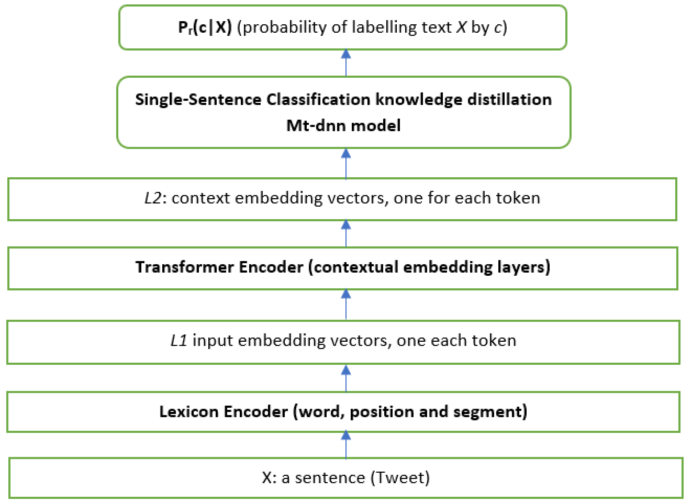


**Fig. 2.** MT-DNN model for representation learning.

of embedding vectors, one for each word, in $l_1$. Then the Transformer encoder captures the contextual information for each word and produces the shared contextual embedding vectors in $l_2$ ($l_2$ is a layer above $l_1$). Finally, the additional task-specific layers generate task-specific representations (single sentence classification task in our case), followed by the process of knowledge distillation. The logistic regression with softmax predicts the probability that tweet(X) is labelled as class c.

**Finding Relevant Terms Set** In the following paragraph, we describe the design of our proposed system framework, as shown in Figure 1. Given a set of microblog corpus. We first perform data cleaning and indexing into the database. A query term will then be matched with the tweet index to retrieve initial result set. The query and initial results are then classified and only those results, which are in the query class, used to expand the initial query with more relevant and frequent terms. The logical flow of the process is detailed in Algorithm 1.

### 3.2 Phase B

Phase (**B**) consists of the following steps: (v) Obtain the a sentence-level vector for the user query by utilising the crisisNLP pretrained word embedding via word2vec method. (vi) Compute the similarity between the query vector obtained in step (v) compute m keyword vectors obtained in step (iv). Select the expansion words with the highest similarity as the query expansion words. (vi) Use the top similar keywords to expand the user query using query expansion technique.

# 4 Experiments

We conducted experiments for finding the best classification model to classify crisis related tweets as micro-text documents and then we used the best model to set up experiments for our query expansion method[4].

We used crisis NLP labelled data sets contain approximately 50k labelled tweets and consist of various event types such as earthquakes, floods, typhoons, etc.

The tweets are initially labelled into various informative classes (e.g., urgent needs, donation offers, infrastructure damage, dead or injured people) and one not-related or irrelevant class. The objective of the experiment was to find out the model that can further and best classify the tweets into event and action classes to have a vocabulary depicting respectively emerging situations (events like water and shelter shortage) and performed actions during a disaster (relief like water delivered to a given area, several rooms available for families).

## 4.1 Data Pre-processing

Pre-processing was required before using Tweets to address issues that characterize them and thereby produce a clean dataset.

*Data cleaning.* In our experiment, we considered that tweet texts are brief, irregular expressions, noisy, unstructured, and often containing misspellings and grammatical mistakes with words out of the dictionary. We removed blank rows, changed all the text to lowercase, removed URLs, re-tweets and user-mentions. Then we moved towards tokenization that broke each tweet in the corpus into a bag of words. Followed by removal of English stopwords, non-numeric and special characters and perform word-stemming/lemmatization. WordNetLemmatizer required pos tags to understand if the word is a noun or verb or adjective (by default, it was set to "noun").

*Indexing data collections.* As a result of indexing the cleaned tweets collection, we created an inverted index matrix that represents the content of the collection. An inverted index is a dictionary where each word is associated with a list of document identifiers in which that word appears. It enables agile access to the position within a document in which a term appears. Indeed, this structure allows avoiding making quadratic the running time of token comparisons. So, instead of comparing, record by record, each token to every other token to see if they match, the inverted indices are used to look up records that match on a particular token.

*Word Embedding Initialisation.* In our CNN experiment setting, we have used crisis embedding to initialise the embedding at the beginning of the experiment. Crisis embedding is a 300-dimensions domain-specific embedding created by [13]

[4] https://github.com/MehrdadFarokhnejad/Classifying_Tweeter_Crisis_Related_Data

trained on 20 million crisis-related tweets using the Skip-gram model of the *word2vec* tool from a large corpus of disaster-related tweets. The corpus contains 57, 908 tweets and 9.4 million tokens.

### 4.2 Applying Classification Techniques

We performed our experiments using three models with the following hyper-parameters initialisation.
**a. Machine Learning Model:** Chosen parameter values for SVM - regularization were (C)=100, kernel type='rbf', gamma value=0.1. For Random Forest-max depth=5, n estimators=10, max features=1.
The label encodes the target variable - This is done to transform Categorical data of string type in the data set into numerical values. Next step is word vectorization by using TF-IDF Vectorizer - This is done to find how important a word in a document is in comparison to the corpus. After fitting the data, we run the machine learning algorithm to check accuracy.
**b. CNN Model.** Values of parameters: Filters no. = 250, Pool size = 2, Hidden size = 128, Kernel size = 3.
We have re-implemented the CNN and Crisis embedding model from [14] to compare it with the other models. We used a multilayer perceptron with a CNN.
**c. MT-DNN Model.** Values of parameters: Learning rate=5e-5, global gradient clipping = 1.0, Learning gamma = 0.1, epoch = 30, Variable batch sizes = (16,32,64).
We applied the latest Microsoft MT-DNN [11] model on our data set looking for better classification performance.

### 4.3 Data Sets and Labels

We have used the CrisisNLP data set for our classification task and measuring the accuracy of all the three models. We modified the labels of tweets such that each crisis event data set has two labels - Event (a situation produced during a disaster) and Action (represents reactions to events). Table 1 shows the number of tweets for each set. The class "Event" includes tweets which subject is related to any occurrence or incidence happening during or after the crisis. For example, "damage happened to a building" or "people are trapped in buildings in downtown". For "Action" we consider those tweets that focus on operations taking place during or after the crisis. Such as government or NGOs providing help to the affected people.

We performed a set of experiments on California and Nepal earthquake, Typhoon Hagupit and Pakistan Flood data sets (see Table 2). The distribution of data is shown in table2: train (70), validation (10) and test sets (20). Column Labels show the total number of annotations for each class.

### 4.4 Classification Results

Table 3 reports the performance of the five models applied to the California, Nepal, Hagupit and Pakistan crisis datasets. Note that for a given model, dataset

**Table 1.** Description of the classes in the data sets.

| Class | Total label | Description |
|---|---|---|
| Event | 1869 | Tweets reporting occurrence and happening of events during the crisis. Reports deaths, injuries, missing, found, or displaced people,infrastructure and utilities damage |
| Action | 2684 | Tweets reporting responses and measures taken by people during crisis.Messages containing donations or volunteering offers also sympathy-emotional support |

**Table 2.** Class distribution of events under consideration and all other crises.

| Class | Nepal | California | Typhoon | Pakistan |
|---|---|---|---|---|
| Event | 688 | 574 | 271 | 356 |
| Action | 1535 | 255 | 462 | 432 |
| Total | 2203 | 829 | 733 | 788 |

quality across different disaster events is not similar and also tweets were noisy and after cleaning there can be some data loss which could affect the contextual understanding of models. Hence the models learn and generalise better events with higher data quality and uniform class label distribution. Note that among all the machine learning models (i.e., SVM, RF and NB), the accuracy score of SVM is comparatively higher than RF and NB. For example, the accuracy score of the California dataset using SVM is **88.17**, whereas the accuracy score for RF and NB are **87.36** and **86.56**, respectively. For the CNN model with crisis embedding, the accuracy scores for the California dataset is **90.13** and for Nepal **88.62**. This is also true for the Typhoon Hagupit and Pakistan Flood datasets, with SVM core accuracy of **87.61** and **92.82**, respectively.

CNN model outperforms the machine learning models in terms of accuracy score. For the MT-DNN model the accuracy scores were: for California crisis= **91.72** and for Nepal crisis= **90.31**. For Typhoon Hagupit= **91.22** and for Pakistan Flood= **95.72**. We can see that the MT-DNN model surpasses the machine learning and the CNN model and have the best accuracy score.

Several works have done crisis tweets classification research [7, 14]. Existing work has addressed different labels using several classes applying RF, linear regression (LR), SVM and CNN. The best accuracy results were obtained with CNN on the Nepal and California Earthquakes, datasets the typhon Hagupit and the cyclone PAM (resp. 86,89, 81.21, 87,83 and 94,17). Our binary classification leads to acceptable accuracy results ranging from 87.36, 86.95, 85.43 and 90.64.

Our results in table 3 show that MT-DNN model performs better than CNN and ML models used in previous experiments, however from computing com-

**Table 3.** Accuracy score of SVMs, RF, NB, CNN and MT-DNN with respect to crisis tweet data.

| Data sets | SVM | NB | RF | CNN | MT-DNN |
|---|---|---|---|---|---|
| California | 88.17 | 86.56 | 87.36 | 90.13 | **91.72** |
| Nepal | 87.21 | 83.83 | 86.95 | 88.62 | **90.31** |
| Hagupit | 87.61 | 80.75 | 85.43 | 89.31 | **91.22** |
| Pakistan | 92.82 | 91.25 | 90.64 | 93.34 | **95.72** |

plexity point of view, since CNN performance was very close to MT-DNN we can also use CNN instead of MT-DNN which will save computation power and cost.

### 4.5 Models Compared: Classification vs Non-classification

In this section, we have presented an ablation study in which we have compared the performance of our proposed classification based query expanding method against the traditional query expanding method. We use the available crisisNLP pre-trained word embedding via word2vec method [7] to obtained query and expansion terms vectors. In the vector space model, all queries and terms are represented as vectors in dimensional space 300. Documents similarity is determined by computing the similarity of their content vector. To obtain a query vector, we represent keywords in user queries as vectors, and then sum all the keyword vectors followed by averaging them. For our analysis, we calculated the average similarity between the query vector and 'm' keyword vectors obtained for a given query by using the formula 1.

$$Similarity(Candidate\_terms, Query) = \frac{\sum_{i=1}^{m}(Cosine(Query\_vector, Term\_vector[i]))}{m} \quad (1)$$

where 'm' is a hyper-parameter in query expansion-based retrieval, which shows the number of expansion terms (ET). Using as reference the studies[23, 3], we set the number of expansion terms to 10, 20 and 30 (ET@10, ET@20, ET@30). We repeat this task for 100 queries and report the mean of average of each **ET@** set in table 4. The experimental results show that the expanded query terms obtained from the classified query expansion model are more similar and relevant than the non-classification model. The ET@10, ET@20 and ET@30 scores of our proposed classification model surpassed the transition non-classification based model. Also, we observe that when we set the number of expansion terms to 10, we achieve the best performance.

## 5 Conclusions and Future Work

This paper introduced a classification based query expansion method. For classification purpose, various machine learning algorithms studied and were validated

**Table 4.** The mean average of Cosine Similarity (MACS)between query and expanded query terms with and without classification model.

| Query Expansion Model | ET@10 | ET@20 | ET@30 |
|---|---|---|---|
| Classification | **0.420** | **0.377** | **0.371** |
| Non-classification | 0.401 | 0.366 | 0.369 |

through experiments using crisis tweet datasets that compared the performance of the applied models. Also, we showed that query expansion base on classification method obtains better candidate expansion words, which are semantically close to the user query. We are currently exploring more robust and advanced NLP models for processing and analyzing crisis data to improve the achieved results. Our future work includes the use of classified vocabularies for exploring data collections using different techniques like queries as answers, query morphing and query by example.

## References


1. Abberley, D., Kirby, D., Renals, S., Robinson, T.: The thisl broadcast news retrieval system. (1999)
2. Acar, A., Muraki, Y.: Twitter for crisis communication: lessons learned from japan's tsunami disaster. International Journal of Web Based Communities **7**(3), 392–402 (2011)
3. Azad, H.K., Deepak, A.: Query expansion techniques for information retrieval: a survey. Information Processing & Management **56**(5), 1698–1735 (2019)
4. Burel, G., Alani, H.: Crisis event extraction service (crees)-automatic detection and classification of crisis-related content on social media (2018)
5. Cameron, M.A., Power, R., Robinson, B., Yin, J.: Emergency situation awareness from twitter for crisis management. In: Proceedings of the 21st International Conference on World Wide Web. pp. 695–698. ACM (2012)
6. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. Journal of machine learning research **12**(Aug), 2493–2537 (2011)
7. Imran, M., Mitra, P., Castillo, C.: Twitter as a lifeline: Human-annotated twitter corpora for nlp of crisis-related messages. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). European Language Resources Association (ELRA), Paris, France (may 2016)
8. Imran, M., Mitra, P., Srivastava, J.: Cross-language domain adaptation for classifying crisis-related short messages. arXiv preprint arXiv:1602.05388 (2016)
9. Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882 (2014)
10. Li, L., Xu, G., Yang, Z., Dolog, P., Zhang, Y., Kitsuregawa, M.: An efficient approach to suggesting topically related web queries using hidden topic model. World Wide Web **16**(3), 273–297 (2013)
11. Liu, X., He, P., Chen, W., Gao, J.: Multi-task deep neural networks for natural language understanding. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 4487–4496. Association for Computational Linguistics, Florence, Italy (Jul 2019). https://doi.org/10.18653/v1/P19-1441, https://www.aclweb.org/anthology/P19-1441

12. Miyanishi, T., Seki, K., Uehara, K.: Improving pseudo-relevance feedback via tweet selection. In: Proceedings of the 22nd ACM international conference on Information & Knowledge Management. pp. 439–448 (2013)
13. Nguyen, D.T., Joty, S., Imran, M., Sajjad, H., Mitra, P.: Applications of online deep learning for crisis response using social media information. arXiv preprint arXiv:1610.01030 (2016)
14. Nguyen, D.T., Mannai, K.A.A., Joty, S., Sajjad, H., Imran, M., Mitra, P.: Rapid classification of crisis-related data on social networks using convolutional neural networks. arXiv preprint arXiv:1608.03902 (2016)
15. Palen, L., Vieweg, S.: The emergence of online widescale interaction in unexpected events: assistance, alliance & retreat. In: Proceedings of the 2008 ACM conference on Computer supported cooperative work. pp. 117–126. ACM (2008)
16. Priya, S., Bhanu, M., Dandapat, S.K., Ghosh, K., Chandra, J.: Taqe: tweet retrieval-based infrastructure damage assessment during disasters. IEEE transactions on computational social systems **7**(2), 389–403 (2020)
17. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World wide web. pp. 851–860. ACM (2010)
18. Spink, A., Wolfram, D., Jansen, M.B., Saracevic, T.: Searching the web: The public and their queries. Journal of the American society for information science and technology **52**(3), 226–234 (2001)
19. Varga, I., Sano, M., Torisawa, K., Hashimoto, C., Ohtake, K., Kawai, T., Oh, J.H., De Saeger, S.: Aid is out there: Looking for help from tweets during a large scale disaster. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 1619–1629 (2013)
20. Vieweg, S., Hughes, A.L., Starbird, K., Palen, L.: Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In: Proceedings of the SIGCHI conference on human factors in computing systems. pp. 1079–1088. ACM (2010)
21. Wang, Y., Huang, H., Feng, C.: Query expansion based on a feedback concept model for microblog retrieval. In: Proceedings of the 26th International Conference on World Wide Web. pp. 559–568 (2017)
22. Xu, B., Lin, H., Lin, Y., Xu, K., Wang, L., Gao, J.: Incorporating semantic word representations into query expansion for microblog information retrieval. Information Technology and Control **48**(4), 626–636 (2019)
23. Zhai, C., Lafferty, J.: Model-based feedback in the language modeling approach to information retrieval. In: Proceedings of the tenth international conference on Information and knowledge management. pp. 403–410 (2001)