Discovering Novel Emergency Events in Text Streams

© Dmitriy Deviatkin[1] © Artem Shelmanov[1] © Daniil Larionov[2]
devyatkin@isa.ru shelmanov@isa.ru dslarionov@protonmail.com
[1]Federal Research Center "Computer Science and Control" of Russian Academy of Sciences,
Moscow, Russia
[2]People's Friendship University of Russia,
Moscow, Russia

Abstract. We present text processing framework for discovering emergency related events via analysis of information sources such as social networks. The framework performs focused crawling of messages, text parsing, information extraction, detection of messages related to emergencies, as well as automatic novel event discovering and matching them across different information sources. For detection of emergency-related messages, we use CNN and word embeddings. For discovering novel events and matching them across different sources, we propose a multimodal topic model enriched with spatial information and a method based on Jensen–Shannon divergence. The components of the framework are experimentally evaluated on Twitter and Facebook data.

Keywords: event detection, topic modelling, monitoring, named entity recognition, text processing, novel topic.

1 Introduction

Recent research showed that Twitter, Facebook, and other social networks have valuable applications in emergency situations. Since large-scale emergency events give rise to a massive publication activity in social networks [35], these resources accumulate information about situation in affected areas, infrastructure damage, casualties, requests and proposals for help. They have already been used for enhancing situation awareness of affected people and emergency response teams [3, 21, 15], as well as for online detecting and monitoring emergency events like earthquakes [27, 29]. Advanced information retrieval techniques can detect emergencies in text streams automatically so direct appeals to the rescue services through the standard channels may not be needed.

This research continues the previous studies presented in [10, 11] that are devoted to monitoring restricted geographical regions via social networks for enhancing situation awareness during emergency situations. In this work, we solve the task of automatic identification of emergency events in a stream of text messages. We consider an event in a text stream as a group of topically related messages that reflect a real-life event in a small time period. Since we are looking for emergency events, it is crucial to detect them as soon as possible: long before they become trendy and gain high amount of publications. Therefore, one of the peculiarities of this task is the problem of identification of novel topics that correspond to emergency events. It is also important to distinguish events (earthquakes, fire breakouts, storms, hurricanes, etc.) that happen in different locations at the same time despite they generate topically similar text streams (e.g. destructions caused by a single storm that moves across a country should be identified as different events).

The task set in this work has a global spatial restriction. In particular, we are interested primarily in the events and messages from the Arctic zone. This restriction brings additional difficulties due to sparseness of data, lack of ready-to-use software, methods, and linguistic resources needed for text processing.

In this work, we evaluate several models for detection of emergency related messages based on various types of embeddings and classification techniques including deep learning. We present a multimodal topic model for event discovering that leverages spatial information, as well as describe approaches to assessing event novelty and matching events from different information sources. The experimental evaluations on collections of messages from Twitter and Facebook show that our methods outperform the baselines.

The rest of the paper is structured as follows. Section 2 reviews the related work on methods for novel topic/event detection in text streams. Section 3 describes the natural language pipeline of our system including the subsystem for extraction of emergency related messages. Section 4 presents the developed method for novel emergency event discovering and matching across information sources. Section 5 describes the experimental evaluation of methods. Section 6 concludes and outlines the future work.

Proceedings of the XX International Conference "Data Analytics and Management in Data Intensive Domains" (DAMDID/RCDL'2018), Moscow, Russia, October 9-12, 2018

2 Related Work

The work related to our current research includes publications considering the tasks of event detection in microblogs, topic evolution tracking, as well as emerging topic detection. Most of the approaches to these problems can be divided into two major groups.

The first group of methods for emerging event detection and tracking primarily relies on topic models adopted to temporal aspects of the task. They are based on different modifications of PLSA models [13] (often LDA [6]). One of the fundamental works in this area is [5]. It proposes several dynamic topic models that align topics across time steps with logistic normal distribution, train with approximation based on variational Kalman filters and perform inference with the help of wavelet regression. Another fundamental model named "topics over time" is presented in [30]. Authors propose a method for jointly modelling both word co-occurrences and localization in continuous time without employing Markov assumption. Another topic model that takes into account temporal dimension is on-line LDA presented in [1]. In this approach, distributions generated on the previous time steps are used as priors for word generation on the current step. For each topic, the method builds transformation matrix that captures the evolution of the topic over time. Authors consider a topic as emerging if it is significantly different from topics in the same time period or from all topics seen before. For topic comparison, Kullback-Leibler divergence is used. In [31], researchers instead of creating monolith Bayesian model propose to learn a topic model and a transition matrix to shift distributions over discrete time steps. They formulate the problem of model learning as minimizing the least square error between predicted topic distribution using transformation and the actual topic distribution of new documents. The proposed approach provides the ability to predict topic trends in the future. Other notable related work on topic models for emerging topic detection in microblog data include Twitter-LDA [12], BBTM (bursty biterm topic model) [34], and TopicSketch [33].

The second group of methods is based on detection of emerging features like terms, keywords, or token segments, and clustering of them. In [7], to define emerging terms authors use two metrics named "nutrition" and "energy function" (biology metaphor). Nutrition of a term is calculated as a sum of modified term frequency in a tweet multiplied by author importance (calculated via PageRank) summed through all tweets in a time period. The energy function of a term is proportional to the difference of its current nutrition and its nutrition in the previous time intervals. Authors declare a term as emerging if its energy value is more than "critical drop" value, which is proportional to the average energy of all terms in the current time period. Using cooccurrence of terms, authors build a graph with edges that correspond to the strongest relationships between terms. The emerging terms become seeds of strongly connected components that finally represent emerging topics. Authors of [32] use wavelet analysis for detection of emerging keywords. They consider frequencies of words as signals and decode these signals with wavelet analysis. Some trivial words are filtered away by analyzing their corresponding signal autocorrelations. The remaining words are then clustered to form events with a modularity-based graph partitioning technique. In [8], a real-time framework for detecting hot emerging topics for organizations in social media context is presented. Authors discover emerging topics and extract emerging features from both the organization and topic perspectives. They extract emerging terms by leveraging chi-square test for foreground and background distributions of terms. Topics are discovered by incremental k-means type clustering algorithm. To perform timely identification of hot emerging topics, authors proposed two semi-supervised classifiers (based on co-training and self-learning). Authors engineered several features that incorporate an authority of a source, importance of keywords, number of retweets, and some other aspects. In [28], the emerging keywords are identified using significance measure based on outlier detection algorithm. More specifically, authors used exponentially weighted average of terms and co-occurring terms. For detection of novel events, in [20], researchers propose to use instead of single unigrams so called "event segments" – key phrases for an event that possibly refer to named entities or semantically meaningful information units. They cluster event segments into events considering both their frequency distribution and content similarity. Emerging segments are detected by abnormal frequency distribution of the tweet and user frequencies of the segments. Importance of an event is also determined by Wikipedia. Authors consider segments that frequently appear as anchors in Wikipedia more favorable. This approach is intended for finding the most realistic events and to derive the most newsworthy segments to describe the identified events.

The method presented in [14] combines two aforementioned approaches: it uses topic modelling in conjunction with models for emerging terms detection. Topic models are used to detect topic distributions in each time interval. Term novelty is estimated by local weighted linear regression. In order to advance from detection of term novelty to detection of topic novelty, authors solve optimization problem. The solution gives novelty and fading probabilities for a topic. Based on these two probabilities, topic evolution operations are defined subsequently to identify emerging topics from the large number of latent ones and track how these topics evolve over time. To compare topics, authors use Jensen-Shannon distance.

Another approach to emergency event detection employ dictionary learning method [17]. The dictionary contains topics, which are consist of atoms (numerical vectors). Vector representation of documents can be approximated with a linear combination of such atoms. The method consists of two steps: determining novel documents in a text stream and identifying a cluster structure among the novel documents. In the first step, the method checks whether a new document can be represented as a sparse linear combination of known atoms with low error. If it is not the case, the document is considered novel. Such documents are used to learn a new dictionary of novel topics. On the second step, the learned dictionary is used to build clusters of similar novel messages. These clusters are considered as emerging topics.

Our approach to novel event discovering is based on multimodal topic modeling and takes into account spatial information. Its key benefits compared to the previous work are the following.

- It allows to separate similar emergency events happened in different locations (for example, storms or typhoons).
- It provides an obvious way to match messages from different sources (social networks) taking into account location information.
- It can help to reveal location information of an event from a set of scattered messages.

3 Natural Language Processing Pipeline

Our method for event discovering needs complex preprocessing of natural language texts. We perform basic linguistic analysis, named entity recognition, time recognition, and detection of emergency related texts.

The final results of the natural language processing pipeline are used for three tasks: focused crawling, enriching information about events, creating modularities for topic models.

3.1 Basic Linguistic Analysis

The basic linguistic analysis includes tokenization, sentence splitting, pos-tagging, lemmatization, and syntax parsing. The pipeline is implemented via IsaNLP[19] – a library that organizes various NLP components for English and Russian. In this paper, we perform experiments only with English texts, therefore, the constructed pipeline contains only components for parsing English.

Tokenization, sentence splitting, postagging, and lemmatization are performed by components based on NLTK toolkit [4]. The syntax parsing is performed by SyntaxNet McParseface [2].

3.2 Named Entity Recognition

We perform extraction of the following types of objects: person's names, organizations, geographical locations, and ship names. For basic NER extraction, we use Polyglot framework. This system uses distant supervision on Wikipedia for learning underlying model and is able to perform named entity recognition for 40 languages. However, we note that performance of such an approach is not suitable for location extraction due to lack of recall. High recall of spatial information is needed to perform filtering of the text stream and topic modelling. Wikipedia lacks many miscellaneous locations, therefore, there is not enough data for training a good model. Polyglot also lacks the ability to normalize locations.

To improve the recall of location extraction and achieve the ability to normalize extracted textual information into geographic coordinates, in the previous work, we implemented a rule- and dictionary-based module [10]. We created a gazetteer from Geonames[20] and supplied it with several filtering rules based on postags of extracted tokens. Geonames also provides mapping of locations into the geographic coordinates.

To extract and normalize temporal expressions, we use a combination of two tools: spaCy[21] (NLP framework based on deep learning) and a datetimeparser[22] (a library based on a set of hand-crafted rules).

For extraction of ship names, in the previous work [11], we implemented a hybrid approach. On the basis of a database of ship names, we implemented a gazetteer that has high recall but low precision due to the fact that many generic words appear to be ship names. To mitigate this problem, we also trained a neural network based on C-LSTM architecture [36]. The network filters out erroneous cases generated by the gazetteer and drastically improves precision and overall F_1-score of ship name detection.

3.3 Detection of Emergency Related Messages

For detection of emergency related tweets, in the previous work, we also used a combination of a gazetteer and a neural network based on C-LSTM architecture. The gazetteer is based on the CrisisLex lexicon, proposed in [23]. This gazetteer generates many false positives that are filtered out by the neural network. To create this solution, in the previous work, we collected a corpus of tweets and trained a neural network on it. In this work, we improve the module for detection of emergency related messages by incorporating more labeled data from CrisisLex corpora [24] and by exploring:

- Various embeddings: word-level: fastText [16] (trained on our own corpus / pre-trained on English Wikipedia), GloVe [26] (Common Crawl with dimension 300 / Twitter with dimension 200), Word2Vec [22], sentence-level: InferSent [9].
- Various types of models: logistic regression (from scikit-learn), random forest (from scikit-learn), gradient boosting on decision trees (LigthGBM algorithm [18]), fully-connected network (FCN), convolutional neural network (CNN), and C-LSTM as before.

[19] https://github.com/IINemo/isanlp
[20] http://www.geonames.org/
[21] https://spacy.io/
[22] https://github.com/scrapinghub/dateparser

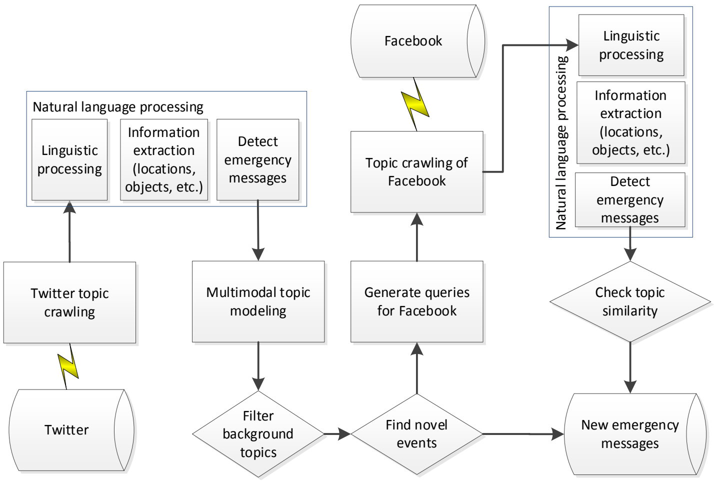

Figure 1. Emergency event detection process

For logistic regression, random forest, gradient boosting algorithms, as well as for FCN we averaged word embeddings and used the result vector as features. Word-level embeddings in C-LSTM and CNN were processed in a standard way. Sentence-level embeddings were not used in C-LSTM and CNN since these architectures work only with sequences. For the rest algorithms, sentence-level embeddings were used as common features.

The fully-connected network is a simple 2-layer perceptron with dropout in the middle. The first layer activation function is ReLU, the outputs of the last layer are passed through the softmax. The architecture of convolutional neural network for sentence classification was proposed in [19]. In this architecture, padded sequence of word embeddings is processed by a one-dimensional convolution layer, followed by max pooling layer to reduce dimensionality. The result vectors are stacked into a single one and are fed into fully-connected layer to make a prediction. Activation functions for convolutional and fully-connected layers are set to ReLU and softmax respectively. The architecture of C-LSTM consists of 1-d convolution layer with ReLU activation and max pooling followed by a LSTM recurrent layer. The final predictions are made by two dense layers with hyperbolic tangent and softmax activations. Neural networks were implemented with PyTorch [25].

4 Emergency Event Detection Method

The pipeline for emergency event detection is depicted in Figure 1. In the first step, we collect all messages from Twitter using topic search API [11] and crisis-related lexicon. Then, we detect emergency related messages among crawled tweets using methods described in section 3.3 and filter out all irrelevant tweets.

In the second step, we train multimodal topic model to identify emergency events described by messages and then determine novel events among them by comparing term distributions of the events from adjacent time periods.

In the third step, we use event-related and location-related lexis from the obtained topics to crawl messages from other sources (Facebook in particular). Then, we apply emergency detection method again and filter out all irrelevant posts. The trained topic model is used to check whether the remaining messages are topically similar to the events extracted from Twitter.

4.1 Identification of Events

In the first step, we discretize the timeline into small time periods (one day in the experiments). In each time period, multimodal topic model with additive regularization [37, 38] is trained.

Let D be a collection of tweets from a time period, let *Def* be a default modality (regular event-related lexis) and let *Loc* be a modality devoted to location of events. The main reason to use such modalities is to separate similar events happened in different places in one period of time. We consider each message $d \in D$ as a set of tokens, related to those modalities $W = W_{def} \cup W_{loc}$. The goal of the topic modeling is to find factorization for matrix of empirical probabilities for documents and tokens:

$$\hat{p}(w|d) \approx p(w|d) = \sum_{t \in T} p(w|t)p(t|d) = \sum_{t \in T} \varphi_{wt}\theta_{td}, \forall w \in W. \quad (1)$$

This problem could be solved by maximizing the weighted sum of the following log-likelihoods with additive regularizers:

$$L(\Phi, \Theta) = \sum_{\gamma \in \Gamma} \gamma \sum_{d \in D} \sum_{w \in W_\gamma} n_{dw} ln \sum_{t \in T} \varphi_{wt}\theta_{td} + \alpha R_{sp}(\Theta) + \beta R_{sm}(\Phi) + \tau R_{dcorr}(\Phi_{loc}) \rightarrow max_{\Phi,\Theta}. \quad (2)$$

Table 3. Results of the models for emergency-related message detection (F_1-score), %

	Embedding features					
Models	**FstTrain**	**FstWiki**	**GloveCC**	**GloveTwt**	**W2V**	**InferSent**
LogReg	87.4±8.4	82.5±9.2	88.6±5.3	85.1±6.9	88.9±6.7	89.4±4.9
Rnd For.	86.9±9.5	82.3±11.1	87.4±7.4	83.9±10.5	87.4±8.9	89.4±4.9
GBDT	91.7±0.1	89.8±0.1	93.0±0.1	89.8±0.2	92.0±0.2	N/A
FCN	90.9±0.3	89.8±0.1	92.2±0.3	88.0±0.2	91.2±0.3	90.8±0.2
CNN	**94.3±0.3**	93.4±0.3	93.8±0.2	92.7±0.2	92.9±0.2	N/A
CLSTM	92.1±0.2	92.2±0.3	92.2±0.6	91.5±0.5	92.3±0.5	N/A

Here $\gamma \in \Gamma = \{\gamma_{def}, \gamma_{loc}\}$ are weights of the modalities, Φ is a matrix of token probabilities for topics, and Θ is a matrix of topic probabilities for documents. As in [37], we apply smooth-sparse regularizers to achieve smooth term distributions in topics and sparse topic distributions in messages:

$$R_{sm}(\Phi) = \sum_{t \in T} KL(\beta_t || \varphi_{wt}), \quad (3)$$

$$R_{sp}(\Theta) = -\sum_{d \in D} KL(\alpha_d || \theta_{td}), \quad (4)$$

where α_d and β_t are sampled from some predefined distributions.

We apply decorrelation regularizer only for location modality to be able to detect similar events happened in different places at the same time:

$$R_{dcorr}(\Phi_{loc}) = -\sum_{t,s \in T} \sum_{w \in W_{loc}} \varphi_{wt}\varphi_{ws}. \quad (5)$$

We use BigARTM library [39] to train multimodal models. The result is Φ and Θ matrices for each time period. After that, "background" topics with high entropy of token distributions can be filtered.

4.2 Detection of Novel Events

In the second step, we determine whether the extracted events were discussed before. We aggregate several adjacent periods of time to "time windows". Consider we have topics s and t in the same time window. Denote vectors of token distributions for these topics as Φ_t and Φ_s. As in [14], we use Jensen–Shannon divergence between token probabilities for the topics to estimate topic similarity:

$$JSD(\Phi_t || \Phi_s) = \frac{1}{2}\big(KL(\Phi_t || M) + KL(\Phi_s || M)\big), \quad (6)$$
$$M = \frac{1}{2}(\Phi_t + \Phi_s).$$

A topic is denoted as a "new event" if there is no earlier similar topics in a predefined time window.

4.3 Events Matching

In the third step, we match messages related to the same event from different sources, which can be various types of social networks or mass media sites. In experiments, we enriched messages from Twitter related to novel emergency events with Facebook public posts. For each novel event, we construct a search query as a combination of default and location tokens with the highest weights. To crawl Facebook, we use Ghost.py[23] library.

We filter obtained posts (leaving only emergency related messages) as described in section 3.3 and extract named entities and locations from them. We infer topic-probabilities matrix $\tilde{\Theta}$ for remaining posts using the pretrained model for the event. Then, we filter all messages, which are not topically similar to the event. Due to the use of multimodal models, information about locations is also taken into account when assessing the similarity of posts.

5 Experiments

5.1 Detection of Emergency Related Messages

Dataset and Pre-processing

For evaluation of method for detection of emergency related messages, we use the CrisisLexT6 dataset. The dataset consists of 60,000 tweets related to 6 major crisis situations. Emergency related tweets are labeled as "on-topic" and others are labeled as "off-topic". The pre-processing procedure included elimination of the special characters, as well as conversion of hashtags, emojis, and URLs into single tokens.

Hyperparameters

Logistic regression. Regularization: L2 penalty. Tolerance: 0.0001. Inverse regularization strength: 1.0.

Random Forest. Number of estimators: 1,000. No limits to maximum number of features and tree depth. Split quality measure: Gini impurity. Min number of samples per split: 2. Min number of samples per leaf: 1.

Gradient boosting. Maximum tree depth: 20. Number

[23] https://github.com/jeanphix/Ghost.py

of leaves: 11. Learning rate: 0.05. Feature fraction 0.9. Bagging fraction: 0.8. Min frequency: 5. Number of estimators: 4,000 with early stopping for 200.

FCN. Size of hidden layer: 256. Dropout: 0.5. Number of epochs: 10. Loss: cross entropy. Optimization algorithm: Adam. Learning rate: 0.0001. Weight decay: 0. Batch size: 256.

CNN. Kernel size: [3, 4, 5]. Number of filters: 512. Dropout: 0.5. Optimization algorithm: Adam. Learning rate: 0.0001. Loss: binary cross entropy. Batch size: 128. Vocabulary size: 10,001. Number of epochs: 10 with early stopping for 3 epochs.

Results and Discussion

We use 5-fold cross-validation for evaluation. Results are presented in Table 1. We discovered several insights into problems with processing and analyzing crisis and Twitter specific lexicon:

- Sentence-level embeddings are better than averaging word vectors. Averaging embeddings of all words in a tweet blur the real meaning of text. InferSent embedding model, which is constructed using NLI data and BiLSTM encoders, treats sentence as a single entity and performs more general projection process. But the higher dimensionality (required to make accurate projections) makes it harder to use several classification algorithms.
- GloVe embeddings pretrained on a Common Crawl corpus show better results than Twitter specific embeddings. Sentence-level embeddings, pretrained on non-specific natural language inference data, also show superior results. It seems reasonable that crisis-related lexicon differs from common Twitter lexicon and tends to be closer to common lexicon. However, we should note that there is a lack of publicly available Twitter data for training. GloVe Twitter corpus contains only 27 billion words, which is much less compared to Common Crawl corpus size of 840 billion words.
- All neural network models have lower standard deviation of F_1-score compared to other machine learning algorithms (except GBDT). Therefore, the quality of neural networks could be much stable on unseen data and less sensitive to the context.
- Our best classifier (CNN for text classification + fastText, trained on our dataset) outperforms models presented in the related work [40, 41, 42].

5.2 Novel Emergency Event Extraction

Dataset and Pre-processing

We crawled 60k Twitter messages from April 1, 2018 to April 12, 2018 using the focused crawler presented in [11]. With the help of CNN neural network, we filtered out messages that are not related to emergency events, which reduced the number of tweets in the dataset to 5,200. The remaining tweets were analyzed with the natural language processing pipeline and with the event discovery method. After that, we also crawled Facebook posts for each extracted event. Using the developed method, we filtered out posts that were considered irrelevant to events extracted from Twitter. After filtering, 1k Facebook posts left.

Hyperparameters

In our experiments, we applied grid search to tune weights of the regularizers for topic models. A criterion for the search was a weighted sum of model perplexity, model's matrices sparsity and model's pointwise mutual information.

Results and Discussion

Since the experiments were conducted on open data, we estimated only precision of models. The results are presented in Table 2. The experiment shows that the proposed approach outperforms baseline LDA models. This confirms the importance of using information about the locations in the framework. One can note relatively low precision for the events matching. We believe this is due to substantial lag of time between the message crawling and the event matching experiments. Thus, true event-related posts may be treated by Facebook's search as less actual than others.

Table 2. Results of the novel emergency event extraction method (Precision), %

Step	*LDA (baseline)*	*Multimodal model*
All events	63.3	93.3
Novel events	71.4	80.0
Event matching	60.0	67.0

6 Conclusion

We considered several problems related to monitoring of social networks: detection of messages related to emergencies, extraction of novel events, and matching events reflected in different text sources. For detection of emergency-related messages, we use CNN and word embeddings. For extraction of novel events and matching them across different sources, we propose a multimodal topic modelling enriched with spatial information and Jensen–Shannon divergence.

We investigated the performance of different algorithms and embeddings for emergency-related message detection on CrisisLexT6 dataset and found that the best solution is given by CNN with fastText embeddings. We also compared the proposed multimodal topic model and the LDA baseline. The experimental results are promising and show that the proposed framework could be useful for monitoring emergency events via messages in social media.

In the future work, we are going to address the problem of emergency event locating and create visualization tools for presenting them on a geographic map.

Acknowledgments. The project is supported by the Russian Foundation for Basic Research, project numbers: 15-29-06082, 15-29-06045 “ofi_m”.

References

[1] Loulwah AlSumait, Daniel Barbará, and Carlotta Domeniconi. On-line LDA: Adaptive topic models for mining text streams with applications to topic detection and tracking. In *Data Mining, 2008. ICDM’08. Eighth IEEE International Conference on*, pages 3–12, 2008.

[2] Daniel Andor, Chris Alberti, David Weiss, Aliaksei Severyn, Alessandro Presta, Kuzman Ganchev, Slav Petrov, and Michael Collins. Globally normalized transition-based neural networks. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 2442–2452, 2016.

[3] Zahra Ashktorab, Christopher Brown, Manojit Nandi, and Aron Culotta. Tweedr: Mining Twitter to inform disaster response. *Proceedings of ISCRAM*, pages 354–358, 2014.

[4] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit*, 2009.

[5] David M Blei and John D Lafferty. Dynamic topic models. In *Proceedings of the 23rd international conference on Machine learning*, pages 113–120, 2006.

[6] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, pages 993–1022, 2003.

[7] Mario Cataldi, Luigi Di Caro, and Claudio Schifanella. Emerging topic detection on Twitter based on temporal and social terms evaluation. In *Proceedings of the tenth international workshop on multimedia data mining*, 2010.

[8] Yan Chen, Hadi Amiri, Zhoujun Li, and Tat-Seng Chua. Emerging topic detection for organizations from microblogs. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 43–52, 2013.

[9] Alexis Conneau, Douwe Kiela, Holger Schwenk, Loc Barrault, and Antoine Bordes. Supervised learning of universal sentence representations from natural language inference data. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 670–680, 2017.

[10] D. Deviatkin and A. Shelmanov. Towards text processing system for emergency event detection in the Arctic zone. In *Proceedings of Data Analytics and Management in Data Intensive Domains*, pages 225–232, 2016.

[11] D. Devyatkin and A. Shelmanov. Text processing framework for emergency event detection in the Arctic zone. *Communications in Computer and Information Science*, pages 74–88, 2017.

[12] Qiming Diao, Jing Jiang, Feida Zhu, and Ee-Peng Lim. Finding bursty topics from microblogs. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 536–544, 2012.

[13] Thomas Hofmann. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 50–57, 1999.

[14] Jiajia Huang, Min Peng, Hua Wang, Jinli Cao, Wang Gao, and Xiuzhen Zhang. A probabilistic method for emerging topic tracking in microblog stream. *World Wide Web*, pages 325–350, 2017.

[15] Muhammad Imran, Carlos Castillo, Ji Lucas, Patrick Meier, and Sarah Vieweg. AIDR: Artificial intelligence for disaster response. In *Proceedings of the companion publication of the 23rd International Conference on World Wide Web Companion*, pages 159–162, 2014.

[16] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. Bag of tricks for efficient text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 427–431, 2017.

[17] Shiva Prasad Kasiviswanathan, Prem Melville, Arindam Banerjee, and Vikas Sindhwani. Emerging topic detection using dictionary learning. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 745–754, 2011.

[18] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, pages 3149–3157, 2017.

[19] Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, 2014.

[20] Chenliang Li, Aixin Sun, and Anwitaman Datta. Twevent: segment-based event detection from tweets. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 155–164, 2012.

[21] Alan M. MacEachren, Anuj Jaiswal, Anthony C. Robinson, Scott Pezanowski, Alexander Savelyev, Prasenjit Mitra, Xiao Zhang, and Justine Blanford. SensePlace2: GeoTwitter analytics support for situational awareness. In *Proceedings of Visual Analytics Science and Technology (VAST) on IEEE Conference*, pages 181–190, 2011.

[22] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.

[23] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. CrisisLex: A lexicon for collecting and filtering microblogged communications in crises. In *Proceedings of ICWSM*, 2014.

[24] Alexandra Olteanu, Sarah Vieweg, and Carlos Castillo. What to expect when the unexpected happens: Social media communications across crises. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 994–1009, 2015.

[25] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in PyTorch. In *NIPS-W*, 2017.

[26] Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[27] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World Wide Web*, pages 851–860, 2010.

[28] Erich Schubert, Michael Weiler, and Hans-Peter Kriegel. Signitrend: scalable detection of emerging topics in textual streams by hashed significance thresholds. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 871–880, 2014.

[29] Sayan Unankard, Xue Li, and Mohamed A Sharaf. Emerging event detection in social networks with location sensitivity. *World Wide Web*, pages 1393–1417, 2015.

[30] Xuerui Wang and Andrew McCallum. Topics over time: a non-markov continuous-time model of topical trends. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 424–433, 2006.

[31] Yu Wang, Eugene Agichtein, and Michele Benzi. Tm-LDA: efficient online modeling of latent topic transitions in social media. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 123–131, 2012.

[32] Jianshu Weng and Bu-Sung Lee. Event detection in Twitter. *ICWSM*, pages 401–408, 2011.

[33] Wei Xie, Feida Zhu, Jing Jiang, Ee-Peng Lim, and Ke Wang. Topicsketch: Real-time bursty topic detection from Twitter. *IEEE Transactions on Knowledge and Data Engineering*, pages 2216–2229, 2016.

[34] Xiaohui Yan, Jiafeng Guo, Yanyan Lan, Jun Xu, and Xueqi Cheng. A probabilistic model for bursty topic discovery in microblogs. In *AAAI*, pages 353–359, 2015.

[35] Jie Yin, Sarvnaz Karimi, Bella Robinson, and Mark Cameron. ESA: emergency situation awareness via microbloggers. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, pages 2701–2703, 2012.

[36] Chunting Zhou, Chonglin Sun, Zhiyuan Liu, and Francis Lau. A C-LSTM neural network for text classification. *arXiv preprint arXiv:1511.08630*, 2015.

[37] Anastasia Ianina, Lev Golitsyn, and Konstantin Vorontsov. Multi-objective topic modeling for exploratory search in tech news. *Conference on Artificial Intelligence and Natural Language,* pages 181–193, 2017.

[38] Konstantin Vorontsov, and Anna Potapenko. Additive regularization of topic models. *Machine Learning*, pages 303–323, 2015.

[39] Konstantin Vorontsov et al. BigARTM: Open source library for regularized multimodal topic modeling of large collections. *International Conference on Analysis of Images, Social Networks and Texts*, pages 370–381, 2015.

[40] Roy Chowdhury S, Purohit H, Imran M. D-sieve: a novel data processing engine for efficient handling of crises-related social messages. *InProceedings of the 24th International Conference on World Wide Web*, pages 1227–1232, 2015.

[41] Zhang S, Vucetic S. Semi-supervised discovery of informative tweets during the emerging disasters. arXiv preprint arXiv:1610.03750. 2016.

[42] Li H, Caragea D, Caragea C, Herndon N. Disaster response aided by tweet classification with a domain adaptation approach. Journal of Contingencies and Crisis Management, pages 16-27, 2018.