Automatic Information Retrieval from Tweets: A Semantic Clustering Approach

Julien Coche, Aurelie Montarnal, Andrea Tapia, Frederick Benaben

Automatic Information Retrieval from Tweets: A Semantic Clustering Approach

Julien Coche
IMT Mines Albi
julien.coche@mines-albi.fr

Aurelie Montarnal
IMT Mines Albi
aurelie.montarnal@mines-albi.fr

Andrea Tapia
Penn State University
axh50@psu.edu

Frederick Benaben
IMT Mines Albi
benaben@mines-albi.fr

ABSTRACT

Much has been said about the value of social media messages for emergency services. The new uses related to these platforms bring users to share information, otherwise unknown in crisis events. Thus, many studies have been performed in order to identify tweets relating to a crisis event or to classify these tweets according to certain categories. However, determining the relevant information contained in the messages collected remains the responsibility of the emergency services. In this article, we introduce the issue of classifying the information contained in the messages. To do so, we use classes such as those used by the operators in the call centers. Particularly we show that this problem is related to named entities recognition on tweets. We then explain that a semi-supervised approach might be beneficial, as the volume of data to perform this task is low. In a second part, we present some of the challenges raised by this problematic and different ways to answer it. Finally, we explore one of them and its possible outcomes.

Keywords

Information Retrieval, Word Embedding, BERT.

INTRODUCTION

Social networks have become a part of many people's daily lives. Smartphones and an ever-increasing Internet allows every citizen to document and report what is happening around them at all times. This behavior is not limited to everyday events but is also observed during natural and man-made disasters. These informal reports have since been considered by emergency services (Cameron et al. 2012; Terpstra and Stronkman 2012). In response to this new trend they decided to develop capabilities to process this new flow of information. However, this new format comes with new requirements.

This capability has been studied for a long time, and a lot of work has been done on the automated processing of messages posted on social networks. As the flow of data is very important on social networks, a significant part of this work is focused on the identification and classification of tweets related to an ongoing event. This has led to the development of a wide range of systems designed to assist emergency operators in processing social media data. Systems such as AIDR (Imran et al. 2014) allow the identification of tweets that are relevant to the operations of emergency services. It also allows to classify these tweets according to their content, among other features. Yet, these tools seem incomplete in terms of process automation, as they still require the intervention of a human operator to watch and process the filtered flow. (Kropczynski et al. 2018) highlights the practices of operators and the needs they have concerning information retrieval during a phone call. These needs are expressed through the 6W's used by American call centers. Thus, when they receive a phone call, operators systematically seek to answer *Where, Why, When, Who, What, Weapon* questions. These 6 questions can be considered as the 6 categories of interest for emergency services. This observation identifies therefore a gap in existing systems, as none of them address this need, as far as we know. In order to meet this need, such system should be able to automatically

retrieve the answers to these questions from social network data. Previous work mentioned earlier would be used as a first filter that would reduce noise from the global stream. The question that the following article will then address is: *How to identify among the messages posted on social networks mentioning an ongoing event, the entities usually looked for by emergency services operators.*

Table 1. Examples of information found in tweets posted by eyewitnesses (CrisisLexT26)

calgary$_{where}$ flooding$_{what}$ again. you have got to be kidding me
There is a growing fire$_{what}$ near Norwest Hospital, NSW$_{where}$. Fire brigade$_{who}$ is in attendance.
Just got out work$_{when}$ in the heart of Sydney CBD$_{where}$ and the smell of smoke$_{what}$ is strong. Eerie sky. Fire$_{what}$ must be huge.

This article is therefore focused on the automation of the retrieval of information contained in tweets posted by witnesses of the event. An example of the research objective is shown Table 1. This task of assigning labels to words contained in a text belongs to the field of natural language processing and is defined as named entity recognition (NER). A lot of work has been done to perform this task at relevant confidence levels. Also, the most efficient methods use models that rely on deep neural networks. However, these models require a lot of training data to obtain adequate performance. An example of benchmark for NER is CoNLL (Kim 2003). This data set contains about 20k sentences for training, validation and testing and aimed at identifying four different types of named entities: persons, locations, organizations and others. On this specific dataset, state of the art deep neural networks models achieve human like performances. However, there is no data set with an equivalent volume (20k sentences labelled at the word level) in crisis informatic. But there are different data sets of past crisis events that can be used to answer our problematic. Ideally, the data set should also contain messages from eyewitnesses mentioning information relevant to the emergency services.

The largest data set of tweets posted during crisis situations is CrisisLex (Olteanu et al. 2014). Several crises have been recorded an made available to help crisis informatic research. Some of these data sets have been labelled. However, CrisisLex has been constructed and labelled with the aim of classifying the tweets at the tweet level, namely, whether they are event-related or not, or according to different categories considered of interest. In order to address the above-mentioned problematic, the data set requires a) data from eyewitnesses of the event, b) labelled data, c) that the labelling is done at the word level. Considering all these elements, no CrisisLex event can be used unless it requires relabeling. Some data sets meet criteria a) and b). However, this represents around 2000 tweets, or 10% of CrisisLexT26 (Olteanu, Vieweg, and Castillo 2015). This volume of data thus appears to be less than what is usually used for training deep neural networks. Moreover, these data are not labelled in a way that allows direct training. In conclusion, as far as we know, there is no data set to answer this problem. So, there are two possibilities: a) create a new dataset, b) use an approach requiring less data. The rest of this article takes the path of proposal b) and presents a method that relies on existing unlabeled data to classify the content of tweets.

RELATED WORK

Our contribution builds on previous work on the following: (1) content-based classification of tweets to determine what a tweet refers to in order to help rescue services filter the content of social networks. (2) Methods to represent the semantics of words in a vectoral way, in order to perform semantics-related processing. (3) Finally, methods that allow NER to be performed in a semi-supervised manner, with interest in methods using label propagation.

Tweet classification in crisis management

A substantial amount of work has been carried out to simplify the processing of data from social networks by emergency services. Several classifications of the information available on social networks have been proposed. Bottom-up approaches, like (Vieweg et al. 2010), use tweets collected during an event to identify different categories of interest. Namely, in this article the categories were: *warning*, *preparatory activity*, *fire line/hazard location*, *flood level*, *weather*, *wind*, *visibility*, *road conditions*, *advice*, *evacuation information*, *volunteer information*, *animal management* and *damage/injury* reports. This approach leads to categories that are more specific to the observed events, reducing the generalization possibilities. Symmetrically, top-down approaches like (Bénaben et al. 2016) build on general, somehow abstract, categories that are defined based on their final use. Because if this, categories could be less specific to an event, allowing a better generalization. From these classifications have emerged different platforms for the identification of relevant data, but also their classification

according to the categories previously proposed (Imran et al. 2014). For a time, these platforms mainly used supervised machine learning algorithms such as Support Vector Machine (SVM) or Decision Trees, which are used to perform the identification and classification tasks mentioned above (Caragea et al. 2011). Later, deep neural network models were trained to perform the same tasks, with a significant gain in performances. (Caragea, Silvescu, and Tapia 2016) use for example an architecture based on a convolutional neural network and note a performance gain on the results of the classifications.

With this classification task comes also other challenges, which can be found in the literature. One of them is the generalization of the models created. Since each crisis is by nature unique, it is difficult to train a single model capable of learning from past crises to generalize to future crises. Research has been done toward transfer learning approaches that would allow to reuse the model learnt from one type of crisis to others without additional training nor data (H. Li et al. 2015). On the other hand, others have attempted to focus on the similarities between each of the crises such as victims, damage, emergency services involved, etc. to build their models (Coche et al. 2019).

However, in order to answer the problematic mentioned in the introduction, the tweet classification is not enough. On the contrary, one seeks to assign labels to entities belonging to defined categories. So, NER is more interesting than tweet classification. (Kim 2003) describes how this task is solved by different systems proposed on a dataset called CoNLL 2003. The entities concerned here are in this case: LOCATION, PERSON, ORGANIZATION and OTHER. Later (Ritter et al. 2011) tried to perform NER on tweets. To do so, they first look for what are the categories of interest in tweets. They identified 10 different categories of interest in a dataset of tweets. They found out that Twitter posts are a challenging format to perform NER, as 140 characters are very short sentences that carry few context (Twitter now allow 280 characters). (C. Li et al. 2012) presented TwiNER, a system able to identify named entities in a Twitter stream. This semi-supervised system was designed to detect emerging crisis using dynamic programming algorithms. It achieved an overall 0.4 on the F1 score on the dataset that they. However, despite being promising, these performances were insufficient for use by emergency services. (Ashktorab et al. 2014) proposed a supervised machine learning approach to address this problem. Using Conditional Random Fields, in a supervised way, they were able to reach 0.6 F1 score in average. Consequently, there is still room for improvements to perform NER on tweets in a semi-supervised to reach supervised approaches performances and beyond.

Language modelling and word semantic representation

In natural language processing, the ability to represent textual data digitally is important. The naivest and used for a long time has been the encoding of words through one hot encoders in such a way that each word in a sentence is represented by its position in its vocabulary. This representation is used in (H. Li et al. 2015) as an input representation in their domain adaptation approach. However, a shortcoming of this representation is that it does not consider the context in which the word is used. Other models were therefore created afterwards.

Context-based, local (Mikolov et al. 2013) and global (Pennington, Socher, and Manning 2014) approaches to the words used in sentence a among the construction of context-sensitive vector representations. This has greatly improved the predictions of the machine learning models used. Note that one of these representations was constructed for the crisis (Imran, Mitra, and Castillo 2016). These representations, in addition to being semantically richer, also provide a much smaller dimensional space than those obtained with one hot encoding. This allows us to maintain a certain coherence between the vectors created and the definition of the closest neighbors of a vector.

This vector representation model was then improved with (Peters et al. 2018) which proposes to consider the different contexts in which the words are seen. Later (Devlin et al. 2018) proposed BERT (Bidirectional Transformers for Language Understanding), obtained from a larger corpus of texts. Its architecture, based on Transformers (Vaswani et al. 2017) instead of the usual Recurrent Neural Networks, allows to parallelize the training of these very large models. Evolutions, using the same architecture have followed, improving the vector representation or reducing the resources needed for their operation (Y. Liu et al. 2019; Sanh et al. 2020).

Label propagation in semi supervised learning for named entities recognition

The machine learning used in natural language processing is mainly based on supervised training models, which rely on labelled training datasets. However, when few labelled data are available, but a large amount of unlabeled data is, it becomes interesting to use semi-supervised models instead. These models use the few labelled data provided and then continue their learning process on the unlabeled data.

This semi-supervised training approach applied to natural language processing (Liang 2005) and NER leads to the development of new models capable of taking advantage of both labelled and unlabeled data. (Liao and Veeramachaneni 2009) use a conditional random fields algorithm trained on the labeled data. Then the algorithm

is used to generate new features from the unlabeled data, that are then used to train the classifier. Later, (X. Liu et al., n.d.) combined K-nearest neighbors and conditional random fields to recognize named entities in tweets, showing improvements in performances using K-nearest neighbors and semi-supervised learning.

Similar to the K-nearest neighbors' approach, label propagation (Zhu and Ghahramani, n.d.) could be used to propagate the labels assigned to a small set of labelled entities, in a semantic graph. This approach shown promising results in named entities recognition in Vietnamese text (Le et al. 2013).

CLASSIFICATION THROUGH WORD VECTOR REPRESENTATION CLUSTERING

In this article we present a model to perform word labelling of social media data according to predefine classes. Our model relies on semantic similarities of words and deep neural networks models that generate word vector representations of words contained in a vocabulary. The resulting matrices (or vector spaces) are of dimension $m \times n$, with m the number of words and n the dimension of the vectors created. In these vector spaces, words that are semantically close have vector representations that are close to each other in distance. So, the overall idea is to use 2 vocabularies. The first one, the *source* data set is composed of text data collected during crisis events and that are representative of the classes being researched. The classes used here are the categories mentioned in (Coche et al. 2019). These categories represent the information that the emergency services are looking for (namely, the *6W's*). This data set is then composed of labeled entities. The second one, the *target* data set is composed of words that are representative of the messages sent during the event. This dataset is not labeled. From these 2 data sets, we extract their respective vocabularies (all the unique words that both datasets are composed).

Then, we define the word vector representation of each word present in both vocabularies. To do so, we use the BERT model (Devlin et al. 2018). For each previously created vocabulary, we define the vector representations of their words. The result of this step is 2 matrices of dimension $m_S \times n$ for the source data set and $m_T \times n$ for the target data set.

These 2 matrices are then merged in a vector space of dimension $(m_S + m_T) \times n$. This new vector space contains both labeled word vector representations and unlabeled ones. In order to label the unlabeled words, we propagate the labels to the unlabeled words around the labeled ones. The distance that define the propagation or not of the labels around the labeled words is defined as the *semantic propagation radius*. The radius in which the labels are going to be propagated is defined by the user. This value is set experimentally in the following experiment. The radius allows to define the number of neighbor's candidates for the propagation of labels. The lower the semantic similarity between the source word and the unlabeled words captured, the larger the radius can be considered. Table 2 shows the evolution of the neighbors for the word "police" for 2 different values of the radius. The table is split in 2 columns, the left columns shows the closest neighbors until the radius reach 0.55 and the right column goes until 0.44. Experimentally, we can observe that the closest words in terms of semantic similarity (radius value of 0.55) correspond to the synonyms of the target word. By widening the radius, we can see what is close to the lexical field of the source word.

DATA AND PREPROCESSING

This section presents the datasets used and how they were processed prior to the experiment

Data set

The model presented above uses 2 datasets. One, unlabeled, is derived from messages posted on social networks in a crisis. The second, labelled, is a vocabulary containing words considered to be representative of the classes we are trying to identify. To obtain these two datasets, we used data from CrisisLex **(Olteanu et al. 2014)**. These two datasets are: the CrisisLexT26 for non-labelled vocabulary, and the CrisisLexRec for labelled vocabulary.

Preprocessing

For the *target* data set, we used the CrisisLexT26 (25k labeled tweets and 225k other unlabeled tweets) and for the *source* data set we used the CrisisLexRec composed of 380 words and word pairs. For the target data set, we only considered the events where users were mainly english speakers. consists of lowercasing the tweets. The cleaning of the target data set then, we removed the "RT" and "#" symbols in the tweets, the punctuation symbols, the mentions and the urls, as they don't bring more information in our case. The tweet is then tokenized using the tokenizer provided along BERT. We only keep the tokens that appear more than 5 times and then remove the remaining stopwords.

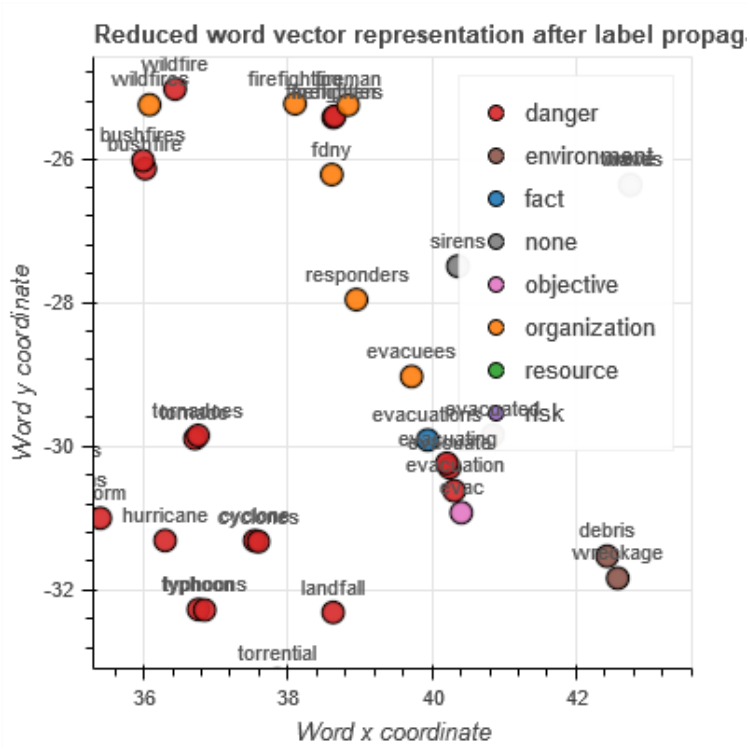

Figure 1. Sample of resulting label propagation after reduction of dimensions by T-SNE (Maaten and Hinton 2008)

For the *source* data set, we used the CrisisLexRec, composed of 380 words or word pairs. We have split the word pairs and only kept one iteration of each word in order to have a vocabulary composed only of unique words. We labeled the data according to concepts described in (Bénaben et al. 2016), *ie.* **resource** (supply, volunteer…), **organization** (policeman, firefighters …), **fact** (destroyed, flooded…), **danger** (fire, flood…), **risk** (alert, threat…), **objective** (reconnect, safe…) and **none**. After labeling the data, we end up with 154 words with a label.

We used the BERT Base multilingual embedding[1] to create the word vector representations. The resulting vectors are of dimension 512. We used the Euclidian distance to compute the semantic similarity between the word vector representations.

EXPERIMENTS AND RESULTS

Our experimental setup is aimed to address the following question: How can we label words in a sentence according to predefine classes?

To address this question, we have implemented the previous methodology applied to a preprocessed version of the CrisisLexT26. In this case, the target data set contains 4183 unlabeled unique tokens and the source data set contains 154 labeled unique tokens. Once the 2 data sets are merged, there are 4028 unique tokens unlabeled.

The label propagation is then performed on the resulting matrix. We have experimentally determined that a value of 0.55 for the semantic propagation radius effectively propagates labels to words close enough semantically. With this value, we propagate the labels to 728 new words. Figure 1 provides an example of the resulting word labelling.

Table 2 shows the evolution of the neighbors for the word "police" for 2 different values of the radius. The table is split in 2 columns, the left columns shows the closest neighbors until the radius reach 0.55 and the right column goes until 0.44. Experimentally, we can observe that the closest words in terms of semantic similarity (radius value of 0.55) correspond to the synonyms of the target word. By widening the radius, we can see what is close to the lexical field of the source word.

[1] https://storage.googleapis.com/bert_models/2018_11_23/multi_cased_L-12_H-768_A-12.zip

Table 2. Neighbors words of the "police" token retrieved for different radius values. The words on the right are consecutives of the words on the left.

R = 0.55	*R = 0.44*	
Cops	...	Po
Cop	Political	Violence
Policy	Arrested	Street
Officer	Government	Cities
Officers	Fbi	Civil
Patrol	Authorities	Acted
Crime	Laws	Depament
Sheriff	Insurance	Weapon
Enforcement	Job	Service
Politics	Govt	Ticket
Law	Feds	Grounds
Military	Offduty	Fireman
Crimes	Pd	Dept
Army	Deputy	Suspect
Arrest	Officials	Agencies
Lapd	Soldier	

CONCLUSION AND FUTURE WORK

This paper presented an approach to address the initial issue: *How to identify the entities usually sought by emergency service operators among the messages posted on social media and mentioning an ongoing event?* This problematic emphasizes the need to identify and retrieve the information contained in the tweets according to a

classification used by the professionals, rather than labeling at the tweet level. While we would normally have used a deep neural network to address this problem, the lack of training data for training the neural network led us to opt for another approach. It consists in labeling words representative of the classes involved, then propagating these labels to words semantically close to the labeled words. After computing the distance between the word vector representation, the labels are propagated if the distance is small enough to be considered as close. Our approach brings benefits:

- It requires fewer labeled data. In our experiment, we only labeled 154 words, while tweet tagging at the word level would have required tagging thousands of words
- It is a non-greedy approach. Only words that are semantically close are labeled, while words that are not related are not labeled.
- It relies mainly on word vector representation models, which allows it to take advantage of future improvements without having to fundamentally change the method.

However, in its current state, our method has certain weaknesses:

- The radius used to define semantically close words must be set by the user and some classes may take advantage of different values for this parameter.
- If a word appears 2 times during the label propagation step (if it is labeled and appears as a semantic neighbor of another word), then only the last label is kept.
- Finally, our approach lack of context awareness. Word labeling is independent of the surroundings words and the general context of the sentence.

The future work will be dedicated to addressing the problems mentioned previously. It also consists in completing the process by implementing the classification of words contained in a tweet and evaluating this method.

REFERENCES

Ashktorab, Zahra, Christopher Brown, Manojit Nandi, and Aron Culotta. 2014. "Tweedr: Mining Twitter to Inform," 5.

Bénaben, F., M. Lauras, S. Truptil, and N. Salatgé. 2016. "A Metamodel for Knowledge Management in Crisis Management." In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, 126–35. https://doi.org/10.1109/HICSS.2016.24.

Cameron, Mark A., Robert Power, Bella Robinson, and Jie Yin. 2012. "Emergency Situation Awareness from Twitter for Crisis Management." In *Proceedings of the 21st International Conference on World Wide Web*, 695–698. WWW '12 Companion. New York, NY, USA: ACM. https://doi.org/10.1145/2187980.2188183.

Caragea, Cornelia, Nathan McNeese, Anuj Jaiswal, Greg Traylor, Hyun-Woo Kim, Prasenjit Mitra, Dinghao Wu, et al. 2011. "Classifying Text Messages for the Haiti Earthquake," 10.

Caragea, Cornelia, Adrian Silvescu, and Andrea H Tapia. 2016. "Identifying Informative Messages in Disaster Events Using Convolutional Neural Networks," 6.

Coche, Julien, Aurélie Montarnal, Andrea Tapia, and Frederick Benaben. 2019. "Actionable Collaborative Common Operational Picture in Crisis Situation: A Comprehensive Architecture Powered with Social Media Data." In *Collaborative Networks and Digital Transformation*, edited by Luis M. Camarinha-Matos, Hamideh Afsarmanesh, and Dario Antonelli, 151–62. IFIP Advances in Information and Communication Technology. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-28464-0_14.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." *ArXiv:1810.04805 [Cs]*, October. http://arxiv.org/abs/1810.04805.

Imran, Muhammad, Carlos Castillo, Ji Lucas, Patrick Meier, and Sarah Vieweg. 2014. "AIDR: Artificial Intelligence for Disaster Response." In *Proceedings of the 23rd International Conference on World Wide Web - WWW '14 Companion*, 159–62. Seoul, Korea: ACM Press. https://doi.org/10.1145/2567948.2577034.

Imran, Muhammad, Prasenjit Mitra, and Carlos Castillo. 2016. "Twitter as a Lifeline: Human-Annotated Twitter Corpora for NLP of Crisis-Related Messages." *ArXiv:1605.05894 [Cs]*, May. http://arxiv.org/abs/1605.05894.

Kim, Erik F Tjong. 2003. "Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition," 6.

Kropczynski, Jess, Julien Coche, Eric Obeysekare, Frederick Bénaben, Rob Grace, Shane Halse, Aurélie Montarnal, and Andrea Tapia. 2018. "Identifying Actionable Information on Social Media for Emergency Dispatch," 11.

Le, Huong Thanh, Rathany Chan Sam, Hoan Cong Nguyen, and Thuy Thanh Nguyen. 2013. "Named Entity Recognition in Vietnamese Text Using Label Propagation." In *2013 International Conference on Soft Computing and Pattern Recognition (SoCPaR)*, 366–70. https://doi.org/10.1109/SOCPAR.2013.7054160.

Li, Chenliang, Jianshu Weng, Qi He, Yuxia Yao, Anwitaman Datta, Aixin Sun, and Bu-Sung Lee. 2012. "TwiNER: Named Entity Recognition in Targeted Twitter Stream." In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 721–730. SIGIR '12. Portland, Oregon, USA: Association for Computing Machinery. https://doi.org/10.1145/2348283.2348380.

Li, Hongmin, Nicolais Guevara, Nic Herndon, Doina Caragea, Kishore Neppalli, Cornelia Caragea, Anna Squicciarini, and Andrea H Tapia. 2015. "Twitter Mining for Disaster Response: A Domain Adaptation Approach," 7.

Liang, Percy. 2005. "Semi-Supervised Learning for Natural Language." Thesis, Massachusetts Institute of Technology. https://dspace.mit.edu/handle/1721.1/33296.

Liao, Wenhui, and Sriharsha Veeramachaneni. 2009. "A Simple Semi-Supervised Algorithm for Named Entity Recognition." In *Proceedings of the NAACL HLT 2009 Workshop on Semi-Supervised Learning for Natural Language Processing - SemiSupLearn '09*, 58–65. Boulder, Colorado: Association for Computational Linguistics. https://doi.org/10.3115/1621829.1621837.

Liu, Xiaohua, Shaodian Zhang, Furu Wei, and Ming Zhou. n.d. "Recognizing Named Entities in Tweets," 9.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. "RoBERTa: A Robustly Optimized BERT Pretraining Approach." *ArXiv:1907.11692 [Cs]*, July. http://arxiv.org/abs/1907.11692.

Maaten, Laurens van der, and Geoffrey Hinton. 2008. "Visualizing Data Using T-SNE." *Journal of Machine Learning Research* 9 (Nov): 2579–2605.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Distributed Representations

of Words and Phrases and Their Compositionality." *ArXiv:1310.4546 [Cs, Stat]*, October. http://arxiv.org/abs/1310.4546.

Olteanu, Alexandra, Carlos Castillo, Fernando Diaz, and Sarah Vieweg. 2014. "CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises." In *Eighth International AAAI Conference on Weblogs and Social Media*. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8091.

Olteanu, Alexandra, Sarah Vieweg, and Carlos Castillo. 2015. "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 - CSCW '15*, 994–1009. Vancouver, BC, Canada: ACM Press. https://doi.org/10.1145/2675133.2675242.

Pennington, Jeffrey, Richard Socher, and Christopher Manning. 2014. "Glove: Global Vectors for Word Representation." In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–43. Doha, Qatar: Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1162.

Peters, Matthew E., Mark Neumann, Luke S. Zettlemoyer, and Wen-tau Yih. 2018. "Dissecting Contextual Word Embeddings: Architecture and Representation." In *EMNLP*.

Ritter, Alan, Sam Clark, Mausam, and Oren Etzioni. 2011. "Named Entity Recognition in Tweets: An Experimental Study." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 1524–1534. EMNLP '11. Stroudsburg, PA, USA: Association for Computational Linguistics. http://dl.acm.org/citation.cfm?id=2145432.2145595.

Sanh, Victor, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2020. "DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter." *ArXiv:1910.01108 [Cs]*, January. http://arxiv.org/abs/1910.01108.

Terpstra, Teun, and R Stronkman. 2012. "Towards a Realtime Twitter Analysis during Crises for Operational Crisis Management," 10.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need." In *Advances in Neural Information Processing Systems 30*, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, 5998–6008. Curran Associates, Inc. http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf.

Vieweg, Sarah, Amanda L. Hughes, Kate Starbird, and Leysia Palen. 2010. "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*, 1079–1088. CHI '10. Atlanta, Georgia, USA: Association for Computing Machinery. https://doi.org/10.1145/1753326.1753486.

Zhu, Xiaojin, and Zoubin Ghahramani. n.d. "Learning from Labeled and Unlabeled Data with Label Propagation," 8.