



Robust Domain Adaptation Approach for Tweet Classification for Crisis Response

Reem ALRashdi^{1,2(✉)} and Simon O’Keefe¹

¹ Computer Science Department, University of York, York, UK
{rmma502, simon.okeefe}@york.ac.uk

² CSSE Department, University of Hail, Hail, Saudi Arabia
r.alreshede@uoh.edu.sa

Abstract. Information posted by people on Twitter during crises can significantly improve crisis response towards reducing human and financial loss. Deep learning algorithms can identify related tweets to reduce information overloaded which prevents humanitarian organizations from using Twitter posts. However, they heavily rely on labeled data which is unavailable for emerging crises. And because each crisis has its own features such as location, occurring time and social media response, current models are known to suffer from generalizing to unseen disaster events when pretrained on past ones. To solve this problem, we propose a domain adaptation approach that makes use of a distant supervision-based framework to label the unlabeled data from emerging crises. Then, pseudo-labeled target data, along with labeled-data from similar past disasters, are used to build the target model. Our results show that our approach can be seen as a general robust method to classify unseen tweets from current events.

Keywords: Domain adaptation · Twitter data · Crisis response · Distant supervision

1 Introduction

Twitter has shown to be one of the most critical sources of gathering information during crises [1] and [2]. Timely information posted by people such as infrastructure damages, injured or dead people, people’s needs and locations can help humanitarian organizations to make quick decisions in real-time, and, hence, saves human lives and reduces financial and economic loss [3]. Unfortunately, information in the people-generated posts cannot be used in daily operations due to information overloaded [4].

Most previous approaches aimed at reducing information overload use supervised learning algorithms to classify tweets into two classes: related and unrelated and heavily rely on manually-labeled data to build accurate models [5] and [6]. However, the lack of manually-labeled data from the current event in real-time prevents the application of such models as it is infeasible to manually annotate tweets for an emerging crisis. Later on, transfer learning has been applied where models are trained on data from past (source) events and used to label tweets from an emerging (target) event [7]. However, [8] and [9] point out that these models cannot successfully generalize to a target event even if the two events come from the same crisis type because

each event has its own characteristics such as location, nature, people response, and infrastructure and economic damages. Thus, models use real-time adaptation techniques to reduce the domain shift between source and target events with the absence of labeled data from the target domain are extremely desirable for crisis response.

Several semi-supervised domain adaptation techniques have been adopted to incorporate unlabeled target data to labeled source data to reduce the gaps between the two domains. According to [10], unlabeled target data can be annotated using pseudo-labeling techniques to be then utilized by retraining a pretrained source model from scratch, finetuning the pretrained model or building a new model. However, to the extent of the author's knowledge, there is no work in studying the application of using distant supervision [11] as a pseudo-labeling technique for tweet classification for crisis response.

Here, we use a distant-supervision-based framework to label the unlabeled target data (pseudo-labeling) where an initial keyword list is established using the available annotated source data from past similar events. The most related keywords are then selected using a statistical method. After that, the chosen keywords list is expanded by employing distant supervision via an external knowledge-base, FrameNet [12], and the tweets having a bigram of keywords are labeled as positive examples while tweets having none of the keywords are labeled as negative examples. Our method is useful when tweets describing the emerging crisis may not include keywords driven from past events as we provide an expanded keyword list via FrameNet. Also, our method avoids the error amplification problem caused by using a basic semi-supervised approach (self-training), especially when the emerging event is different from past events.

Our work is similar to [7], [7] builds an online model where Nepal Earthquake data (target data) are received in batches to finetune a pretrained source model; however, they believe that an amount of labeled target data is available while we assume not. Also, we build a target model by incorporating pseudo-labeled target data into manually-labeled source data instead of pretraining and finetuning a source model.

Our contributions are summarized as follows: (1) We introduce a distant-supervision-based framework that gives pseudo labels to unlabeled target data to be then used with labeled source data to build a robust model to classify unseen tweets from emerging events, (2) we investigate the model performance for crisis-related data and compare it to another pseudo-labeling technique in three adaptation methods and (3) we evaluate our work on eight 2012–2015 crisis events from three crisis types (earthquake, floods, and typhoon).

2 Related Work

Domain adaptation techniques have been successfully applied in many Natural Language Processing (NLP) tasks [13]. Among domain adaptation researches, the most relevant is the works introduce domain adaptation approaches for disaster response to reduce the domain shifts between past and emerging events tweets. [14] proposes the first semi-supervised approach uses labeled source data and unlabeled target data on three classification tasks. Their self-training iterative method shows promising results, mainly when used to classify tweets related to a specific crisis. [15] extends the work

by comparing Naïve Bayes and self-training with hard labels (NB-ST) to Naïve Bayes and Expectation-maximization with soft labels (NB-EM) in classify tweets related to an emerging disaster. The results show that, in general, NB-ST is better than NB-EM when evaluated on the CrisisLexT6 dataset. Also, a hybrid feature-instance adaptation approach has been proposed by [16] to choose a subset of the source crisis data that is similar and can represent the target crisis to build a Naïve Bayes target classifier. The results show that the hybrid approach is better than using feature-based nor instance-based methods individually. Another recent work, [17], extends domain adaptation with adversarial training proposed in [18] to include a graph-based semi-supervised learning introduced by [19]. The unified framework consists of supervised, semi-supervised, and domain adversarial components to learn the similarity between source and target domains and a good domain discriminator. F1 score on only two datasets (Queensland Floods and Nepal Earthquake) improves the performance with 5%–7% absolute gain. Although previous works showed great results towards using domain adaptation for crisis response, there is still room for improvement to reach the performance of supervised target classifiers when labeled target data is available [15].

Recent NLP studies have shown the effectiveness of using distant supervision to label training data [20–23]. [20] employs distant supervision for event extraction using frames from FrameNet as event types and the linguistic units as triggers that evoke the event. [21] proposes a combination framework of a relational and a linguistic knowledge bases on Wikipedia data, Freebase and FrameNet, respectively. Unlike the previous works, we use the lexical features of the available manually-labeled tweets along with an external linguistic knowledge base. In the context of applying distant supervision on Twitter data, several studies have been conducted. [22] applies distant supervision to the topic classification task where they transfer labels from tweets of topically-focused Twitter accounts to tweets posted by general Twitter accounts. [23] uses YouTube videos to assign labels to tweets containing links to these videos. Our work also applies distant supervision on Twitter data; however, we use an external knowledge base and driven lexical features from the existing human-labeled tweets.

According to [24], domain adaptation can be achieved by building a target model using manually-labeled source data with pseudo-labeled target data. We believe that there have been no works on domain adaptation approaches that use distant supervision-based framework to classify disaster-related tweets from an emerging event. Thus, this paper focuses on using a distant supervision-based framework to give unlabeled emerging tweets pseudo labels to be then incorporated to labeled source data from several similar past disasters to build a robust crisis-related classifier and compares it to the widely used pseudo-labeling technique (Self-labeling).

3 Method

Our method (described in Algorithm 1) contains two stages: the pseudo-labeling stage and the adaptation stage. In the pseudo-labeling stage, unlabeled tweets from the current (target) crisis event are gathered using Twitter API. Then, the unlabeled tweets are given pseudo labels by applying our distant supervision-based framework. In the adaptation stage, the pseudo-labeled target tweets are then used to build a target model

with several crisis events from different time intervals and locations from the same crisis type to the given target event.

Algorithm 1: Robust domain adaptation approach with pseudo-labeled target data.

1. Given: labeled tweets of several disaster events from different time intervals and locations from the same crisis type to the given target event (MLS), unlabeled tweets from target domain (UT) retrieved using Twitter API using publicly available tweets ids, and manually-labeled test data from target domain (MLTT)
 2. **Pseudo-labeling stage:** Use our framework to label UT based on all the available MLS and employing distant supervision via the external knowledge base (giving them pseudo labels).
 3. **Adaptation stage:** Build a target model using MLS with the pseudo-labeled data from the target domain.
 4. Evaluate the model on MLTT.
-

3.1 Pseudo-labeling Stage

The distant supervision-based framework used to give pseudo labels to the unlabeled emerging disaster data (in Fig. 1) consists of five stages as follows:

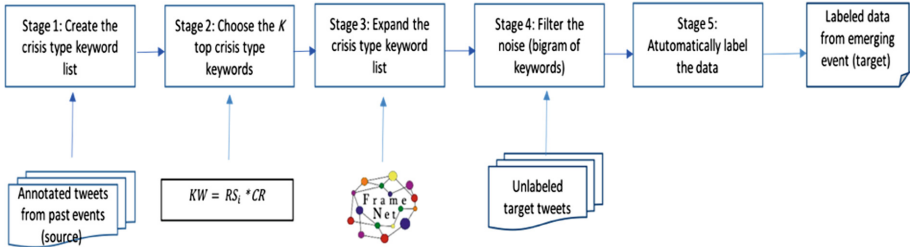


Fig. 1. The distant supervision-based framework.

Stage 1. An initial list of keywords is created based on the available annotated tweets from different events related to the same crisis type. This list includes an unlimited number of words without any restrictions. To avoid word redundancy, we use the Snowball Stemmer tool from NLTK 3.4 to stem each word to its root.

Stage 2. The top K ($K = 10$ in our experiments) keywords from the initial list are selected based on an intrinsic filtering method where we calculate the Keyword (KW) value of each word. In a tweet, a word that describes a given crisis type can be a verb, a noun, or an adverb. For example, magnitude (noun), shake (verb) and deadly (adv) are keywords of the crisis type Earthquake. Intuitively, a word describing a crisis type

appears more than other words in the related tweets. Also, if the same word appears in both related and unrelated tweets, then it has a low probability of being a keyword of this crisis type. Thus, KW is calculated as follows:

$$RS_i = \text{Count}(W_i, CT) / \text{Count}(CT) \quad (1)$$

$$CR_i = \log(3 / (\text{Count}(CTC_i))) \quad (2)$$

$$KW_i = RS_i * CR_i \quad (3)$$

Where RS_i (Role Saliency) represents the saliency of i -th keyword to identify a given crisis type, $\text{Count}(W_iCT)$ is the number of a word W_i occurs in all the tweets related to the crisis type CT and $\text{Count}(CT)$ is the count of times all words occurring in all the crisis-related examples. The KW equation is inspired by [21] where they use a similar Key Rate (KR) value to detect key arguments in event extraction tasks; however, unlike [21], CR_i (Crisis Relevance), in our work, represents the ability of the i -th keyword to distinguish between the related and unrelated tweets to a given crisis type, and $\text{Count}(CTC_i)$ equals 1 if the i -th keyword occurs only in the related tweets and 2 otherwise. Finally, and after removing stop words such as “and”, hashtags such as “#earthquake”, places such as “Nepal” and useless twitter-specific words such as “RT” and “via”, we compute KW_i for all the words in the initial list from stage one and sort them according to their KW values to pick the top K keywords of a given crisis type. For example, for disaster type Earthquake, the words “earthquake,” “hit,” and “magnitude” have the highest KW values comparing to other words in the initial list.

Raw word frequency can be seen as a weak measurement for calculating the importance of word for a specific category due to the skews where words like stop words can be widely-used but not informative. However, we already eliminate this disadvantage by removing all such words and stemming all words to their roots. The imbalanced data problem does not affect our formula as $\text{Count}(CT)$ takes into account the total number of words in the related tweets only while the total number of words in the unrelated tweets is ignored. Other methods, such as Pointwise Mutual Information (PMI) [25] or Term Frequency-Inverse Document Frequency (TF-IDF) [26] have not been used here for solid reasons. PMI, where we calculate the final PMI score of the positive and the negative examples, is not a fair metric in our case because of the imbalanced and limited data problem where the number of positive examples is higher than the number of negative patterns in all the crisis events. On the other hand, our method does not take into consideration the amount of negative examples.

TF-IDF is not a suitable method because IDF has more impact on the final result than TF, where, in our case, they should be equally important since tweets are short and full of noise. If we use TF-IDF on our data, rare words such as misspelled words will have higher TF-IDF than essential keywords. For example, in Earthquake crisis type data, “earthquake” may appear very frequently in related Earthquake event tweets and once or twice in unrelated tweets. On the other hand, our method does not discard the impact of word frequency if the word appears in both related and unrelated tweets.

Stage3. The K top keyword list is expanded to include new similar linguistic units from an external linguistic knowledge-base, FrameNet. FrameNet for English consists of more than 1000 semantic frames. Each frame is associated with a group of Lexical Units (LUs), lemmas and part of speech tags. LUs evoke the frame, which, in our work, are used as crisis keywords that evoke the crisis event. Here, we map each keyword in the keyword list to linguistic units in FrameNet associated with the related frames only.

Stage 4. The unlabeled target data is filtered by using a specific lexical feature (bi-grams of keywords). Only examples with two keywords from the final keyword list remain. This stage eliminates tweets with only one weak keyword (expanded from FrameNet), thus, decreases the noise caused by stage three.

Stage 5. A collection of labeled data from the emerging crisis event is automatically generated by labeling the filtered tweets from stage four as related examples and tweets with no keywords as unrelated examples.

3.2 Adaptation Stage

To classify tweets from unseen disaster events, we build a new target model using the pseudo-labeled target data labeled by our framework and the available manually-labeled source data from the same crisis type. Data generated by our distant supervision-based framework provides new keywords than the keywords driven from source data. Adding these examples to the training data brings target-related features such as location and crisis nature.

4 Experiments

We use two methods to give pseudo labels to the unlabeled target data (stage 1): our distant supervision-based framework (DS), and pretrained model on MLS (SelfL). For the adaptation stage, we use three methods to incorporate target data (stage 2): modifying all the weights in the pretrained model (Finetuning (FT)), fixing all the layers except the output layer (Feature eXtraction (FX)), and building a new model using source and target domain data (TM). To determine the effectiveness of using pseudo-labeled target data generated by our framework, we compare the following eight classifiers (supervised (SL) and semi-supervised (SSL) learning structures) on eight settings (shown in Table 1): (1) SL-LT: trained on MLTT (upper limit), (2) SL-LS: pretrained on MLS (lower limit) (3) SSL- DS-TM, (4) SSL-SelfL-TM, (5) SSL-DS-FX, (6) SSL-SelfL-FX, (7) SSL- DS-FT, and (8) SSL-SelfL-FT. We use the best-reported classifier in [27] for crisis-related tweets. It consists of CNN and Bi-LSTM [28] layers with a pretrained 100-dimensional Glove embedding [29]. All the models are tested on MLTT. In the evaluation process, we use weighted F1 score because of the imbalanced datasets. And due to the stochastic nature of the learning algorithm, we repeated each experiment 30 times, and the average score is reported in Table 3.

Table 1. Source and target sets for each setting (S) on our experiments.

Settings	Source sets	Target set
S1	Earthquake events: 2014-Chile, 2015-Nepal, 2013-Bohol, 2013-Pakistan	2014-California Earthquake
S2	Earthquake events: 2014-California, 2015-Nepal, 2013-Bohol, 2013-Pakistan	2014-Chile Earthquake
S3	Typhoon events: 2015-Pam, 2014-Odile, 2013-Yolanda, 2013-Oklahoma, 2012-Sandy	2014-Hagupit Typhoon
S4	Earthquake events: 2014-Chile, 2014-California, 2013-Bohol, 2013-Pakistan	2015-Nepal Earthquake
S5	Earthquake events: 2014-Chile, 2014-California, 2015-Nepal, 2013-Bohol	2013-Pakistan Earthquake
S6	Typhoon events: 2014-Odile, 2013-Yolanda, 2014-Hagupi, 2013-Oklahoma, 2012-Sandy	2015-Pam Cyclone
S7	Floods events: 2013-Queensland, 2013-Manila, 2013-Colorado, 2014-India, 2014-Alberta	2014-Pakistan Floods
S8	Floods events: 2014-Pakistan, 2013-Manila, 2013-Colorado, 2014-India, 2014-Alberta	2013-Queensland Floods

4.1 Datasets

We use labeled and unlabeled datasets. The labeled is publicly available in three datasets: CrisisNLP [30], CrisisLexT26 [31] and CrisisLexT6 [32]. Details about the available source (past) labeled data are given in Table 2. The unlabeled tweets for the eight target datasets are retrieved by their ids available in CrisisNLP. Source and target datasets are shown in Table 1 for each setting in our experiments. In the pre-processing stage, we clean all input tweets by removing emojis, HTTP addresses, numbers, hashtags, user mentions, NON-ASCII letters, and punctuations. We convert all inputs to lowercase and split them into tokens to be passed to the model.

Table 2. Information about the manually-labeled source data from past events.

Datasets	Collections (crises)	Related tweets	Nonrelated tweets	Total number of tweets
CrisisNLP	Nepal Earthquake	2839	177	3016
	Chile Earthquake	1648	364	2013
	California Earthquake	169	13	182
	Pakistan Earthquake	1676	336	2012
	India Floods	1500	502	2002
	Pakistan Floods	1985	27	2012
	Hagupit Typhoon	1779	233	2012
	Pam Cyclone	1515	497	2012
	Odile Hurricane	178	4	182

(continued)

Table 2. (continued)

Datasets	Collections (crises)	Related tweets	Nonrelated tweets	Total number of tweets
CrisisLexT26	Bohol Earthquake	969	30	999
	Queensland Floods	919	280	1199
	Colorado Floods	924	74	998
	Manila Floods	920	79	999
	Alberta Floods	982	17	999
	Yolanda Tornado	939	108	1047
CrisisLexT6	Sandy Floods	2010	429	1581
	Oklahoma Tornado	2010	241	1769

4.2 Results and Discussion

As can be seen in the first row in Table 3, SL-LS can be helpful when classifying target data especially in settings 1, 4 and 7 where one or more source events and target events are similar in other features rather than crisis type (nearby locations or close occurring time). This outcome is consistent with earlier studies [8, 14, 15]. Although they have different labeling and adaptation methods, SSL-DS-FX and SSL-SelfL-TM have similar results when testing on various target events. This is possible because they both use the same weights of the pretrained source model to label or classify the target data. We also observe that domain adaptation techniques are not always better than supervised learning models learned from only source data. For example, FT (with self-labeled target data) drops the Nepal Earthquake model’s performance by 0.9%, and FX (with both labeling methods) declines the Chile Earthquake model’s performance by 6.9%. This result is not consistent with [15] where iterative domain adaptation techniques are used. The most interesting observation is that incorporating pseudo-labeled target data generated by our distant-supervision-based framework into the training data improves the performance in all the eight datasets (SSL-DS-FT for settings 2 and 5,

Table 3. Results of our experiments in weighted F1 score for eight models in eight settings. The upper limit and the best-reported results are highlighted in bold.

Models/Settings	S1	S2	S3	S4	S5	S6	S7	S8
SL-LS	0.883	0.812	0.841	0.905	0.785	0.702	0.960	0.680
SSL-DS-TM	0.935	0.864	0.879	0.906	0.773	0.779	0.975	0.794
SSL-SelfL-TM	0.892	0.743	0.856	0.907	0.792	0.688	0.971	0.677
SSL-DS-FX	0.890	0.743	0.858	0.907	0.790	0.698	0.973	0.691
SSL-SelfL-FX	0.883	0.743	0.851	0.907	0.788	0.683	0.966	0.680
SSL-DS-FT	0.874	0.871	0.844	0.896	0.802	0.768	0.972	0.750
SSL-SelfL-FT	0.892	0.743	0.853	0.907	0.787	0.692	0.969	0.677
SL-LT	0.886	0.912	0.902	0.915	0.856	0.894	0.972	0.899

SSL-DS-FX for setting 4 and SSL-DS-TM for the five remaining settings). SSL-DS-TM can be seen as the best general approach among the other six classifiers regardless of the similarity between source and target domains as it reports the best results in five out of eight settings and a minimal gap compared to the best score in the others (<3%). This is not the case for the other classifiers as can be seen in rows 4 and 6, where FT is better than FX when one or more source and target events are different. Surprisingly, in settings 1 and 7, our method is better than the upper limit where the supervised model is learned only from the manually-labeled target data.

5 Conclusion and Future Work

In this paper, we provide a simple but powerful semi-supervised domain adaptation approach for tweet classification for crisis response by using a distant supervision-based framework to label the unlabeled target tweets. Our framework provides a new set of keywords rather than the ones driven by the available past events, helping in adding new features to the training data. The experimental results show that our framework is better than using pretrained models trained on source data to label the unlabeled current events in three different adaptation methods. Building a target model using the labeled target domain data generated by the distant supervision-based framework and the available most-related source domain data improves the target classifier performance on seven out of eight datasets- from 0.1% to 11.4% absolute gain in F1 score. This perfectly suits our task because it requires a small time at event onset, and it can be considered to be a general approach without a need to predefine the similarity between source and target domains unlike the other methods. In the future, we plan to use co-training on two models uses DS with different adaptation methods and choose the agreed labels only.

References

1. Qu, Y., Huang, C., Zhang, P., Zhang, J.: Microblogging after a major disaster in China: a case study of the 2010 Yushu earthquake. In: *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, pp. 25–34. ACM, March 2011
2. Starbird, K., Palen, L., Hughes, A.L., Vieweg, S.: Chatter on the red: what hazards threat reveals about the social life of microblogged information. In: *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, pp. 241–250. ACM, February 2010
3. Vieweg, S.E.: *Situational awareness in mass emergency: a behavioural and linguistic analysis of microblogged communications*. Doctoral dissertation, University of Colorado at Boulder (2012)
4. Gao, H., Barbier, G., Goolsby, R.: Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intell. Syst.* **26**(3), 10–14 (2011)
5. Caragea, C., Silvescu, A., Tapia, A.H.: Identifying informative messages in disaster events using convolutional neural networks. In: *International Conference on Information Systems for Crisis Response and Management*, pp. 137–147, May 2016

6. 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](https://arxiv.org/abs/1608.03902) (2016)
7. 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](https://arxiv.org/abs/1610.01030) (2016)
8. Verma, S., Vieweg, S., Corvey, W.J., Palen, L., Martin, J.H., Palmer, M., Anderson, K.M.: Natural language processing to the rescue? Extracting “situational awareness” tweets during mass emergency. In: ICWSM, pp. 385–392, July 2011
9. Tapia, A.H., Moore, K.: Good enough is good enough: overcoming disaster response organizations’ slow social media data adoption. *Comput. Support Coop. Work. (CSCW)* **23** (4–6), 483–512 (2014)
10. Ruder, S.: *Neural Transfer Learning for Natural Language Processing*. National University of Ireland, Galway (2019)
11. Mintz, M., Bills, S., Snow, R., Jurafsky, D.: Distant supervision for relation extraction without labeled data. In: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, vol. 2, pp. 1003–1011. Association for Computational Linguistics, August 2009
12. Baker, C.F., Fillmore, C.J., Lowe, J.B.: The Berkeley framenet project. In: *Proceedings of the 17th International Conference on Computational Linguistics*, vol. 1, pp. 86–90. Association for Computational Linguistics, August 1998
13. Chu, C., Wang, R.: A survey of domain adaptation for neural machine translation. arXiv preprint [arXiv:1806.00258](https://arxiv.org/abs/1806.00258) (2018)
14. Li, H., Guevara, N., Herndon, N., Caragea, D., Neppalli, K., Caragea, C., Neppalli, K., Caragea, C., Squicciarini, A.C., Tapia, A.H.: Twitter mining for disaster response: a domain adaptation approach. In: *ISCRAM*, May 2015
15. Li, H., Caragea, D., Caragea, C., Herndon, N.: Disaster response aided by tweet classification with a domain adaptation approach. *J. Contingencies Crisis Manag.* **26**(1), 16–27 (2018)
16. Mazloom, R.: *Classification of Twitter disaster data using a hybrid feature-instance adaptation approach*. Doctoral dissertation (2018)
17. Alam, F., Joty, S., Imran, M.: Domain adaptation with adversarial training and graph embeddings. arXiv preprint [arXiv:1805.05151](https://arxiv.org/abs/1805.05151) (2018)
18. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V., Lempitsky, V.: Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* **17**(1), 2030–2096 (2016)
19. Yang, Z., Cohen, W.W., Salakhutdinov, R.: Revisiting semi-supervised learning with graph embeddings. arXiv preprint [arXiv:1603.08861](https://arxiv.org/abs/1603.08861) (2016)
20. Chen, Y., Liu, S., Zhang, X., Liu, K., Zhao, J.: Automatically labeled data generation for large scale event extraction. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Long Papers*, vol. 1, pp. 409–419 (2017)
21. Zeng, Y., Feng, Y., Ma, R., Wang, Z., Yan, R., Shi, C., Zhao, D.: Scale up event extraction learning via automatic training data generation. arXiv preprint [arXiv:1712.03665](https://arxiv.org/abs/1712.03665) (2017)
22. Mohammed, S., Ghelani, N., Lin, J.: Distant supervision for topic classification of tweets in curated streams. arXiv preprint [arXiv:1704.06726](https://arxiv.org/abs/1704.06726) (2017)
23. Magdy, W., Sajjad, H., El-Ganainy, T., Sebastiani, F.: Distant supervision for tweet classification using YouTube labels. In: *ICWSM*, pp. 638–641, April 2015
24. Wang, M., Deng, W.: Deep visual domain adaptation: a survey. *Neurocomputing* **312**, 135–153 (2018)
25. Church, K.W., Hanks, P.: Word association norms, mutual information, and lexicography. *Comput. Linguist.* **16**(1), 22–29 (1990)

26. Sparck Jones, K.: A statistical interpretation of term specificity and its application in retrieval. *J. Doc.* **28**(1), 11–21 (1972)
27. Alrashdi, R., O’Keefe, S.: Deep learning and word embedding for tweet classification for crisis response. In: The 3rd National Computing Colleges Conference (NC3). arXiv preprint [arXiv:1903.11024](https://arxiv.org/abs/1903.11024), October 2018
28. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
29. Pennington, J., Socher, R., Manning, C.: Glove: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014)
30. Imran, M., Mitra, P., Castillo, C.: Twitter as a lifeline: human-annotated Twitter corpora for NLP of crisis-related messages. arXiv preprint [arXiv:1605.05894](https://arxiv.org/abs/1605.05894) (2016)
31. Olteanu, A., Vieweg, S., Castillo, C.: What to expect when the unexpected happens: social media communications across crises. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing, pp. 994–1009. ACM, February 2015
32. Olteanu, A., Castillo, C., Diaz, F., Vieweg, S.: CrisisLex: a lexicon for collecting and filtering microblogged communications in crises. In: Proceedings of the AAAI Conference on Weblogs and Social Media (ICWSM 2014). AAAI Press, Ann Arbor (2014)