Robust Domain Adaptation Approach for Tweet Classification for Crisis Response

Abstract. Information posted by people on Twitter during crises can significantly improve crisis response towards reducing human and financial loss. Deep learning algorithm can identify related tweets to reduce information overloaded which prevents humanitarian organizations from using Twitter posts. Yet, 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 an unseen crisis event when pretrained on past events. To overcome this problem, we propose a domain adaptation approach that makes use of distant supervision-based framework to label the unlabeled data from emerging events. Then, pseudo-labeled target data along with labeled-data from similar past events 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 emerging events.

Keywords: domain adaptation, Twitter data, crisis response, distant supervision.

1 Introduction

Twitter has proven to be one of the most important sources of gathering information during crises [1] and [2].Timely information posted by people such as infrastructure damages, injured or dead people, people needs and locations can help humanitarian organizations to take quick decisions in real time, and, hence, saves human lives and reduces financial and economic loss [3]. Unfortunately, information driven from 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: relevant and nonrelevant, 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 event. 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 like earthquake because each event has its own characteristics like 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 where no labeled data from target domain is available are greatly 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 labeled using pseudo-labeling techniques before incorporated with labeled data by retraining a pretrained source model from scratch, finetuning the pretrained model or building a new model. However, to the best of our 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 selected 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 tweets while tweets having none of the keywords are labeled as negative tweets. 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 the 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 assume that a small 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 the method 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 approach uses labeled source data and unlabeled target data on three classification tasks. Their self-training iterative method shows promising results specifically when used to classify tweets related to a specific crisis. [15] extends the work by comparing Naïve Bayes and self-training with hard labels to Naïve Bayes and Expectation-maximization with soft labels in classify tweets related to an emerging crisis. The results show that, in general, NB-ST is better than NB-EM when evaluated on 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 be used to build a Naïve Bayes target classifier. The results show that the hybrid approach is better than using feature-based nor instance-based approaches 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) improvs with 5%-7% absolute gain. Although previous works showed great results towards using domain adaptation for crisis response, there is still a room for improvements 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], [21], [22] and [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. To the best of our knowledge, there has been no works on domain adaptation approaches that uses distant supervision-based framework to classify crisis-related tweets from an emerging event. Thus, this paper focuses on using distant supervision-based framework to give unlabeled emerging tweets pseudo labels to be then incorporated to labeled source data from several similar past events to build a robust crisis-related classifier, and compares it to the widely used pseudo-labeling technique (a pretrained model on source data).

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 crisis 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 external knowledge base (giving them pseudo labels).
3. **Adaptation stage:** Build a target model using MLS with the pseudo-labeled data from 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 event data in pseudo-labeling stage (in Fig. 1) consists of five stages as follows:

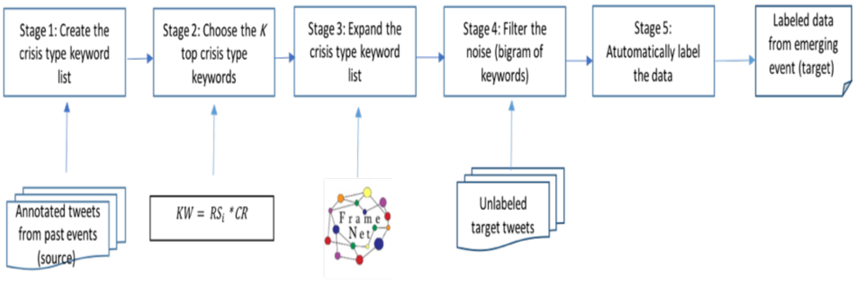
Stage 1: Create the crisis type keyword list
Stage 2: Choose the K top crisis type keywords
Stage 3: Expand the crisis type keyword list
Stage 4: Filter the noise (bigram of keywords)
Stage 5: Autmotatically label the data
Labeled data from emerging event (target)
Annotated tweets from past events (source)
$KW = RS_i * CR$
Frame Net
Unlabeled target tweets

Fig. 1. The distant supervision-based framework.

Stage 1. An initial list of keywords is created based on the available annotated tweets from different event data related to the same crisis type. This list includes unlimited number of words without any restrictions. To avoid word redundancy, we use 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 keyword. 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. In addition, if the same word appears in both related and unrelated tweets, then it has a low probability to be a keyword of this crisis type. Thus, KW is calculated as follows:

$$\mathrm{RS_i} = \mathrm{Count(\,W_i\,,\,CT) \,/\, Count(CT)} \qquad \textbf{(1)}$$

$$\mathrm{CR_i} = \log\,(\,3\,/\,(\,\mathrm{Count(\,CTC_i\,)}\,)\,) \qquad \textbf{(2)}$$

$$KW_i = RS_i * CR_i \quad \textbf{(3)}$$

Where RS_i (Role Saliency) represents the saliency of *i-th* keyword to identify a specific word of a given crisis type, $Count(W_i, CT)$ is the number of a word W_i occurs in all the tweets related to the crisis type CT and $Count(CT)$ is the count of times all words occurring in all the tweets related to the crisis type. 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 tweets related to the crisis type and nonrelevant tweets, and $Count(\ CTC_i\ )$ equals 1 if the *i-th* keyword occurs only in the related tweets and 2 if the *i-th* keyword occurs in both related and nonrelevant tweets. 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 crisis 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 poor measurement for calculating the importance of word for a specific category due to the skews where words like stop words such as the or of can be very frequent 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, where the number of related tweets is more than the number of unrelated tweets, does not affect our formula as $Count(CT)$ 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 PMI for positive examples and PMI for negative examples to calculate the final PMI score, is not a fair metric in our case because of the imbalanced data problem given the limited available manually-labeled data where the number of positive examples is higher than the number of negative examples in all events. On the other hand, our method does not take into consideration the number of negative examples.TF-IDF also is not suitable in our case 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 important keywords. For example, in Earthquake crisis type data, "earthquake" word may appear very frequently in related Earthquake event tweets and once or twice in unrelated Earthquake event 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 similar linguistic units from an external linguistic knowledge-base for English, FrameNet, consists of more than 1000 semantic frames which have more than 100,000 Lexical Units (LU), lemmas and part of speech tags, which in our work are used as crisis keywords. Each frame in FrameNet is associated with a group of LUs that evoke that frame. 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 (bigrams 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 tweets and tweets with no keywords as unrelated tweets.

3.2 Adaptation stage

We add the pseudo-labeled target data to the available labeled source data from the same crisis type of the target crisis to build a new target domain to classify the unseen tweets from the emerging event. Pseudo-labeled target data generated by our distant supervision-based framework provides new keywords than the keywords driven from source data. Adding these data to the training data brings target-related features to the training data 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 the pretrained 100-dimentional 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 set for each settings (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	Floods events: 2013-Queensland, 2013-Manila, 2013-Colorado, 2014-India, 2014-Alberta.	2014-Pakistan Floods
S7	Typhoon events: 2014-Odile, 2013-Yolanda, 2014-Hagupi, 2013-Oklahoma, 2012-Sandy.	2015-Pam Cyclone
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 is 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
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 from 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 event are similar in other features rather than crisis type (nearby locations or close occurring time). This outcome is consistent with earlier studies [8], [14] and [15]. Although they have different labeling and adaptation methods, SSL-DS-FX and SSL-SelfL-TM have similar results when testing on different target events. This is possibly because they both use the same weights of the pretrained source model either 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 Nepal Earthquake model's performance by 0.9% and FX (with both labeling methods) declines 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, 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 very small gap compared to the best score in the others (< 3%). This is not the case 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 supervised model is learned only from the manually-labeled target data.

Table 3. Results of our experiments in weighted F1 score for eight models on 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**

5 Conclusion and future work

In this work, we introduce 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 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. (2011, March). 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.
2. Starbird, K., Palen, L., Hughes, A. L., & Vieweg, S. (2010, February). 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.
3. Vieweg, S. E. (2012). Situational awareness in mass emergency: A behavioural and linguistic analysis of microblogged communications (Doctoral dissertation, University of Colorado at Boulder).
4. Gao, H., Barbier, G., & Goolsby, R. (2011). Harnessing the crowdsourcing power of social media for disaster relief. IEEE Intelligent Systems, 26(3), 10-14.
5. Caragea, C., Silvescu, A., & Tapia, A. H. (2016, May). Identifying informative messages in disaster events using convolutional neural networks. In International Conference on Information Systems for Crisis Response and Management (pp. 137-147).
6. Nguyen, D. T., Mannai, K. A. A., Joty, S., Sajjad, H., Imran, M., & Mitra, P. (2016). Rapid Classification of Crisis-Related Data on Social Networks using Convolutional Neural Networks. arXiv preprint arXiv:1608.03902.
7. Nguyen, D. T., Joty, S., Imran, M., Sajjad, H., & Mitra, P. (2016). Applications of online deep learning for crisis response using social media information. arXiv preprint arXiv:1610.01030.
8. Verma, S., Vieweg, S., Corvey, W. J., Palen, L., Martin, J. H., Palmer, M., ... & Anderson, K. M. (2011, July). Natural Language Processing to the Rescue? Extracting" Situational Awareness" Tweets During Mass Emergency. In ICWSM (pp. 385-392).
9. Tapia, A. H., & Moore, K. (2014). Good enough is good enough: Overcoming disaster response organizations' slow social media data adoption. Computer Supported Cooperative Work (CSCW), 23(4-6), 483-512.
10. Ruder, S. (2019). Neural Transfer Learning for Natural Language Processing. National University of Ireland, Galway.
11. Mintz, M., Bills, S., Snow, R., & Jurafsky, D. (2009, August). 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: Volume 2-Volume 2 (pp. 1003-1011). Association for Computational Linguistics.
12. Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998, August). The berkeley framenet project. In Proceedings of the 17th international conference on Computational linguistics-Volume 1 (pp. 86-90). Association for Computational Linguistics.

13. Chu, C., & Wang, R. (2018). A survey of domain adaptation for neural machine translation. arXiv preprint arXiv:1806.00258.
14. Li, H., Guevara, N., Herndon, N., Caragea, D., Neppalli, K., Caragea, C., ... & Tapia, A. H. (2015, May). Twitter Mining for Disaster Response: A Domain Adaptation Approach. In ISCRAM.
15. Li, H., Caragea, D., Caragea, C., & Herndon, N. (2018). Disaster response aided by tweet classification with a domain adaptation approach. Journal of Contingencies and Crisis Management, 26(1), 16-27.
16. Mazloom, R., Li, H., Caragea, D., Imran, M., & Caragea, C. Classification of Twitter Disaster Data Using a Hybrid Feature-Instance Adaptation Approach.
17. Alam, F., Joty, S., & Imran, M. (2018). Domain Adaptation with Adversarial Training and Graph Embeddings. arXiv preprint arXiv:1805.05151.
18. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., ... & Lempitsky, V. (2016). Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, *17*(1), 2096-2030.
19. Yang, Z., Cohen, W. W., & Salakhutdinov, R. (2016). Revisiting semi-supervised learning with graph embeddings. *arXiv preprint arXiv:1603.08861.*
20. Chen, Y., Liu, S., Zhang, X., Liu, K., & Zhao, J. (2017). Automatically labeled data generation for large scale event extraction. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (Vol. 1, pp. 409-419).
21. Zeng, Y., Feng, Y., Ma, R., Wang, Z., Yan, R., Shi, C., & Zhao, D. (2017). Scale Up Event Extraction Learning via Automatic Training Data Generation. arXiv preprint arXiv:1712.03665.
22. Mohammed, S., Ghelani, N., & Lin, J. (2017). Distant Supervision for Topic Classification of Tweets in Curated Streams. arXiv preprint arXiv:1704.06726.
23. Magdy, W., Sajjad, H., El-Ganainy, T., & Sebastiani, F. (2015, April). Distant Supervision for Tweet Classification Using YouTube Labels. In ICWSM (pp. 638-641).
24. Wang, M., & Deng, W. (2018). Deep visual domain adaptation: A survey. Neurocomputing, 312, 135-153.
25. Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. Computational linguistics, 16(1), 22-29.
26. Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. Journal of documentation, 28(1), 11-21.
27. Alrashdi, R. & O'Keefe, S. (2018, October). Deep learning and Word embedding for tweet classification for crisis response. In The 3rd National Computing Colleges Conference (NC3). arXiv preprint arXiv: 1903.11024
28. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
29. Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543).
30. Imran, M., Mitra, P., & Castillo, C. (2016). Twitter as a lifeline: Human-annotated twitter corpora for NLP of crisis-related messages. arXiv preprint arXiv:1605.05894.
31. Olteanu, A., Vieweg, S., & Castillo, C. (2015, February). 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 (pp. 994-1009). ACM.
32. A. Olteanu, C. Castillo, F. Diaz, S. Vieweg. 2014. CrisisLex: A Lexicon for Collecting and Filtering Microblogged Communications in Crises. In Proceedings of the AAAI Conference on Weblogs and Social Media (ICWSM'14). AAAI Press, Ann Arbor, MI, USA.