Mining Help Intent on Twitter During Disasters via Transfer Learning with Sparse Coding

Bahman Pedrood[1(✉)] and Hemant Purohit[2]

[1] Computer Science Department, George Mason University, Fairfax, VA, USA
bpedrood@gmu.edu
[2] Information Sciences and Technology Department,
George Mason University, Fairfax, VA, USA
hpurohit@gmu.edu

Abstract. Citizens share a variety of information on social media during disasters, including messages with the intentional behavior of seeking or offering help. Timely identification of such help intent can operationally benefit disaster management by aiding the information collection and filtering for response planning. Prior research on intent identification has developed supervised learning methods specific to a disaster using labeled messages from that disaster. However, rapidly acquiring a large set of labeled messages is difficult during a new disaster in order to train a supervised learning classifier. In this paper, we propose a novel transfer learning method for help intent identification on Twitter during a new disaster. This method efficiently transfers the knowledge of intent behavior from the labeled messages of the past disasters using novel Sparse Coding feature representation. Our experiments using Twitter data from four disaster events show the performance gain up to 15% in both F-score and accuracy over the baseline of popular Bag-of-Words representation. The results demonstrate the applicability of our method to assist realtime help intent identification in future disasters.

Keywords: Help seeking · Help offering · Intent · Transfer learning
Crisis

1 Introduction

Social media and Web 2.0 provides a platform for citizens to discuss real-world disaster events, where they share a variety of information including messages to seek help [3]. A recent survey [27] shows the expectation of citizens to get a response from emergency services on social media. As a result, resource-constrained response agencies have started to leverage social media platforms to both communicate and monitor data to enrich their situational awareness for disaster response coordination [10,20,26,32]. Timely extraction of relevant social

Table 1. Examples of messages with intent to seek or offer help during recent disasters.

Social media message	Intent
I wanna give #blood today to help the victims #sandy	*Offering help*
Does anybody know a direct place to send clothing you want to donate to #HurricaneHarvey victims??	*Offering help*
OU students are seeking donations to provide relief and help victims of the Hurricane Harvey disaster: _URL_	*Seeking help*
You can help us send clean water, food, and shelter to those impacted by #HurricaneHarvey! #HoustonStrong	*Seeking help*

media messages allows a better understanding of needs in the affected community. Mining messages with intent to seek or offer help provides specific insights to assist resource coordination by bringing the awareness of actors seeking and offering resources to help. The uncoordinated efforts between such actors and the response organizations can lead to a second disaster of resource mismanagement[1]. Table 1 provides some examples of messages with help intent. However, the intent is not always explicitly expressed in social media messages due to a variety of ways to communicate intentionality. Actionable information can be identified by understanding intent during disasters, likewise, research on modeling intent in user queries on search engines improved the retrieval of relevant results [2].

Our problem of intent identification from social media messages is a form of text classification, however, intentional behavior is focused on future action (e.g., an act of offering donation) in contrast to topic or sentiment/emotion. Prior works on intent classification during disaster events [8,16,21,22,24,34] have developed event-specific supervised learning models using labeled dataset of the corresponding event. However, there are two key limitations of the prior research. First, supervised learning models developed for specific past events do not generalize due to differences in the distributions of the training event data and the testing event data. Second, developing a new supervised learning model onset a future disaster event would require preparing a large labeled dataset quickly to perform well. Therefore, we propose a novel method for the intent classification using transfer learning approach. Transfer learning focuses on leveraging knowledge gained while solving one problem (e.g., identifying intent in a past disaster) and applying it to a different but related problem (e.g., identifying intent in a future disaster). We study four disaster event datasets for the experimental evaluation with an intent class set of {*seeking, offering, none*} and present result analysis for different transfer learning settings of single and multiple disaster data sources. Our specific contributions are the following:

---

[1] https://www.npr.org/2013/01/09/168946170/thanks-but-no-thanks-when-post-disaster-donations-overwhelm.

– We present the first study of realtime intent identification for help-seeking and help-offering messages on social media in a future disaster event by leveraging past event datasets via transfer learning.
– We demonstrate the efficacy of a data-driven representation of Sparse Coding in contrast to the popular Bag-of-Words (BoW) model, to efficiently learn and transfer the knowledge of intentional behavior from past events.
– We evaluate the proposed method in transferring knowledge from both single and multiple past events to a future disaster event, and show performance up to 80% for F-score, indicating good prediction ability in general.

The rest of the paper is organized as follows. We first discuss related work on mining social media for crisis informatics and mining help intent in Sect. 2. We describe our proposed approach in Sect. 3 and experimental setup in Sect. 4. Lastly, we discuss the results and future work directions in Sect. 5.

2   Related Work

In the last two decades, there has been extensive research in the area of crisis informatics for the use of social media in all phases of pre, during, and post disasters (*c.f.* [3,11]). Among different types of social media analytics for disasters, the content-driven analyses are focused on the nature of social media messages including topics such as damage [12,35] and behaviors such as help-seeking [23,24]. User-based analyses include modeling of user attributes, such as trustworthiness [1] and types such as government agencies [17]. Network-based analyses are focused on information diffusion and user engagement patterns, such as communication of official response agencies [33] and retweeting [31].

The proposed research is closely related to content-based analysis of modeling help-seeking behavior on social media. Prior work by [24] identified actionable message category of seeking or offering help while studying Yushu Earthquake in 2010. [21,34] proposed supervised machine learning classifiers to identify and match messages with the complementary intent of request-offer during the events of Great East Japan Earthquake 2011 and Hurricane Sandy 2012 respectively. [22,23] proposed methods for supervised learning and linguistic rule-based classifiers to identify messages with seeking and offering help intent, however, without studying the generalization of methods to leverage the labeled data from past event datasets to mine intent in the future events. [16] proposed a system to classify requests for help during Hurricane Sandy 2012 using n-grams and context-based features. [8] studied the dynamics of messages coordinated by a common hashtag *#PorteOuverte* with the intent to seek or offer help during 2015 Paris Attack and developed an automated classifier for such messages. In the aforementioned works, there was a focus on developing event-specific methods to identify relevant messages with help intent. There is a lack of investigation on how to leverage and transfer the knowledge of intent behavior observed in the past events to help quickly identify relevant messages in the future events and thus, we propose to study transfer learning [18] techniques for mining intent.

3   Approach

We propose a novel approach of transfer learning method for the problem of real-time intent identification in future disasters. This problem is challenging due to differences in the probability distributions of source and target event datasets, the imbalance of intent classes across past and future events, and the lack of effective data representation for inferring intent from the short text.

To address the representation challenge in machine learning and natural language understanding, there is a growing interest in exploring Sparse Coding representation [6,13] in contrast to the popular BoW model. Sparse Coding provides a succinct representation of training text instances using only unlabeled input data, by learning basis functions (e.g., latent themes in the message text) to constitute higher abstract-level features. Given the complexity of expressing intent by multiple combinations of word senses in the text, we hypothesize the use of Sparse Coding to efficiently capture, learn, and transfer intent behavior from past disasters. Next, we describe dataset preparation for past disasters and the features for the proposed model of transfer learning with Sparse Coding.

3.1   Dataset Preparation

This study is based on Twitter messages ('tweets') collected during the past large-scale disasters with a focus on hurricanes and typhoons. We acquired two datasets of tweets annotated for help intent classes of {*seeking, offering, none*} from our past work [22], for two events – Hurricane Sandy 2012 and Supertyphoon Yolanda 2013. We also collected tweet datasets during the recent two disasters in 2017 – Hurricane Harvey and Hurricane Irma, which caused extensive devastation in the United States. We used Twitter Streaming API to collect English language tweets using 'filter/track' method for a given set of keywords (Harvey: {#harvey, hurricane harvey, Harvey2017, HurricaneHarvey, harveyrelief, houstonflood, houstonfloods, houwx}, Irma: {hurricane irma, hurricaneirma, HurricaineIrma, Hurricane Irma, HurricaineIrma, #hurricaneirma2017, #irma, #hurcaneirma}). We collected 8,342,404 tweets from August 29 to September 15 for Hurricane Harvey and 861,503 tweets between September 7 to September 21 for Hurricane Irma. For labeling of help intent classes {*seeking, offering, none*} in each event dataset, we employed a biased sampling approach to increase the coverage of help intent messages given the sparse distribution of intent classes observed in the first two datasets. First, we randomly sampled 2000 tweets from the full dataset of an event and second, we randomly sampled 2000 tweets from the donation-classified subset of messages that provide context for expressing intent. We used the related work's donation topic classifier [21]. We asked three human annotators (no author was involved) to label a tweet into the three exclusive help intent classes for each event and chose the majority voting scheme for finalizing the labels. The resulting labeled class distribution for both the acquired and the collected datasets[2] is shown in Table 2.

---

[2] Datasets are available upon request, for research purposes.

3.2   Feature Representation

Prior work in crisis informatics for social media text classification has extensively used BoW model for representing text messages (*c.f.* survey [11]). However, the BoW representation model limits capturing context and semantics of the text content [5,6], which is essential for inferring intent. Since BoW model loses ordinal information of text content, one direction of research to tackle this challenge is to enhance the text representation in a way that can preserve some ordered information of the words, such as N-grams [19]. Because N-gram representation adds extra terms to the word vocabulary, the problem of curse of dimensionality gets worse. The increase in the feature space, especially for the tasks of small-scale datasets and short text, causes the loss of generalization of the training models and results in a negative effect in the prediction capability.

Table 2. Labeled data distribution for help intent classes across four disaster events, ordered by time of occurrence.

| Event	Month	Seeking	Offering	None
Hurricane Sandy	*2012 (October)*	1626	183	1326
Supertyphoon Yolanda	*2013 (November)*	197	91	475
Hurricane Harvey	*2017 (August)*	816	331	2853
Hurricane Irma	*2017 (September)*	530	244	3226

Sparse Coding is an effective approach for reducing dimensionality of feature space that generally assumes an over-complete basis set for the input data and it is capable of a complete description and reconstruction of the input data. Also, every input data point can be described using a linear combination on a small number of the new basis vectors. According to the theory of compressed sensing, when the data are distributed on an underlying manifold with the characterizing bases, only a small number of bases are required to fully describe any arbitrary point on the manifold [4]. Sparse Coding representation has shown significant improvements in transfer learning in the recent years. While most of the works addressed the challenges of image processing and computer vision [7,9,15,25], there are few research studies for text analytics [14]. However, the previous works primarily required a large amount of data in either source or both source and target domains to be effective, and cannot be applied in our crisis informatics case directly. Because both source and target sets are not only at the small scale but also contain redundant and incomplete information in short text.

For feature extraction, we first perform standard text-preprocessing on the text message to remove stop-words, replace numbers (e.g., money donation mentions), user mentions, and URLs with constant characters as well as lower-casing the tokenized text of messages. After constructing a vocabulary of the extracted tokens, every message is represented by a real-valued vector of vocabulary entries, where each component holds the *tf-idf* value of that entry [30]. After the text samples are transformed to numeric vectors of tf-idf values, a

sparse representation of the vectors in a feature space with a significantly reduced dimension is learned and explained in Sect. 3.3.

3.3   Learning Model: Transfer Learning with Sparse Coding

To tackle the intent mining problem when both source and target data are small, we propose a novel transfer learning approach. Transfer learning has become a popular solution to bring information from past experiences to better characterize the data and class distribution of them. Generally, there exists a large number of labeled data samples from past relevant experiences as well as a large set of unlabeled samples for the new problem, where the two sets are assumed to share common information. The goal of transfer learning is to find better transformations to map the distribution of one of the datasets to the other and learn a predictive model using the transformed data to find a stronger prediction of the new data. Fig. 1 shows an abstract workflow of our proposed method.
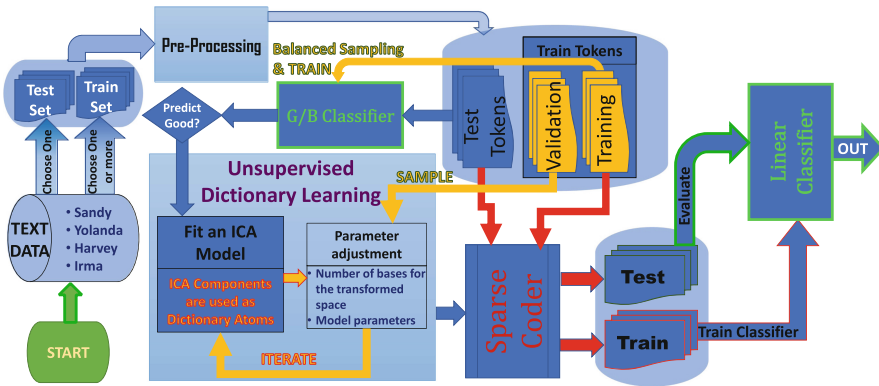
Fig. 1. The sequence of steps in the proposed model. On the top, after pre-processing of train and test data, test set is used for unsupervised learning of *dictionary atoms*. The optimal parameters of the model are learned by cross-validation on the training data. Then a linear classifier is learned from the coded samples in the sparse space.

To formalize the approach, assume a set of message text data, where each data point is coming from a *Domain $\mathcal{D}$*. The given points are usually assumed to be sampled independently and randomly from the domain $\mathcal{D}$. A domain $\mathcal{D}_l = (\mathcal{X}_l, \mathcal{Y}_l)$ is defined as a distribution of pairs of $\{x_{il}, y_{il}\}$, where $\forall \{x_{il}, y_{il}\} \in \mathcal{D}_l : x_{il} \in \mathcal{X}_l, y_{il} \in \mathcal{Y}_l$ and the domain assumes a joint probability distribution $p(\mathcal{X}_l, \mathcal{Y}_l | \theta_l)$. In transfer learning and domain adaptation, we have a labeled *Source* dataset D_S, where data is given as N_S pairs of observations and labels, $\{X_S, Y_S\}$, and those pairs are assumed as i.i.d. samples from the joint probability distribution of the source domain $p(\mathcal{X}_S, \mathcal{Y}_S | \theta_S), \mathcal{D} = \{\mathcal{X}_S, \mathcal{Y}_S\}$. The data of interest is defined as *Target* dataset D_T, where only X_T is given, but Y_T is unknown. The target data points are also assumed as N_T i.i.d. samples from the target domain $\mathcal{D}_T = \{\mathcal{X}_T, \mathcal{Y}_T\}$, which along with their *latent* labels are

jointly distributed as $p(\mathcal{X}_T, \mathcal{Y}_T | \theta_T)$. For a domain $\mathcal{D} = \{\mathcal{X}, \mathcal{Y}\}$, a *task* is defined as a function $f(.)$ on the attribute space $\mathcal{X}$ of that domain to predict the labels $\mathcal{Y}$. In transfer learning, the task is to find a $f(x_{it})$ that approximates y_{it}. In other words, a function $f(x_{it})$ must be learned to predict the values of $\hat{y}_{it}$ for every $x_{it} \in X_T$ that minimizes $\sum_{i=1}^{N_T} d(\hat{y}_{it}, y_{it})$, where $d(.,.)$ is some distance measure of interest. Considering the target domain distribution of $p(\mathcal{X}_T, \mathcal{Y}_T | \theta_T)$, in order to have the most accurate predictions, ideally $f(x_{it})$ must become the closest possible to $p(y_{it} | x_{it}, \theta_T)$. The posterior distribution of target data is unlikely to estimate without having any labels for this data. Since the task in our problem is semantically a complete equivalence in both source and target data, the basic assumption of transfer learning for homogeneous datasets is valid in our context and the joint probability of the labels is assumed to be similar in both source and target domains. Formally:

$$p(X_S) \times p(Y_S | X_S, \theta_S) = p(X_T) \times p(Y_T | X_T, \theta_T) \tag{1}$$

Note that although the joint distribution is assumed as equal for the two datasets, the conditional probabilities are generally not. If the conditional probabilities were equal, then the key difference lies in the marginal distributions. Geometrically, source and target samples are assumed as being scattered in different regions in the feature space. Since $P(X_S) \neq P(X_T)$, the parameters θ_S and θ_T should also be different to compensate the difference of marginal distributions in the conditional probabilities. To fill the gap between the datasets, most of the research in transfer learning is about finding a good transformation $T_{TS}(.) : \mathcal{X}_T \to \mathcal{X}_S$ from the target to source domain, then using a classifier trained on the source data to predict the labels for the transformed target data: $\hat{Y}_T = f(X_T^{ts} | \hat{\theta}_S)$. While this approach works well in some cases, since none of the characteristics of the target data is incorporated in training the classifier, the generalization of the model for the target data reduces. A better approach is a reverse form of transformation: finding a transformation $T_{ST}(.) : \mathcal{X}_S \to \mathcal{X}_T$ from source data to make the marginal distribution of transformed source data similar to the target data.

Our Sparse Coding approach is also about using the unlabeled target data to find a mapping for the source data. The proposed method is distinctive in handling the limitations of intent mining in the specific case of crisis informatics by providing a solution to work with a small number of short-text samples. Instead of only transforming the data, the proposed model combines the domain transfer with the feature reduction step to make the representation generalize well. Fig. 1 shows the steps in the proposed learning model, where it shows how target data is used to build the model, while independent subsets of training data tailor the representation. Sparse Coding in general is about using the characteristics of input data $X \in \mathbb{R}^{N \times M}$ while finding a set of basis $B \in \mathbb{R}^{N \times K}$ that is over-complete on the underlying manifold of data and then, approximating each input instance as a linear combination of a small number of those bases (*atoms*) in the *dictionary*. If the manifold assumptions hold on the data, Sparse Coding guarantees a perfect reconstruction. However, the assumption cannot be easily

confirmed, especially when dealing with a small number of data instances. Also, prior research strongly suggests using an under-complete basis for Sparse Coding in classification [28, 29]. The general form of Sparse Coding is about minimizing $\|X - BA\|_2^2$ while, simultaneously approximating the bases and the sparse code vectors incorporates the following optimization problem:

$$\operatorname*{argmin}_{B, a_i} \frac{1}{2} \left\| \sum_{i=1}^{M} x_i - B a_i \right\|_2^2 - \lambda \|a_i\|_1 \qquad (2)$$

where $\|.\|_p$ is the p-norm. The main objective function that tries to minimize the information loss of coding is a loss function for reconstructing the input sample x_i from the transformed sample a_i, formulated as a Sum of Squared Error (SSE) function. In the regularization term, the parameter λ is introduced to control the sparsity of the coded samples and the *first norm* is used for efficient convergence to a sparse solution. It guarantees a more generalizable solution for the objective function by removing the coded coefficients with an energy level less than λ.

Prior to Dictionary Learning, a logistic classifier is learned on training data in order to improve quality of data for learning the representation, by contrasting the most important intent classes (*seeking, offering*) against the rest (*other* class). Then the dictionary is learned by applying Independent Component Analysis (ICA) method only on the target samples that are classified as relevant on this classifier. To find the best parameter setting for the dictionary, the parameters λ and K (the number of atoms) are selected with a focus on optimizing the resulting metric (F-score) using a $10-$fold Cross-Validation over the important classes *seeking* and *offering* of the training data.
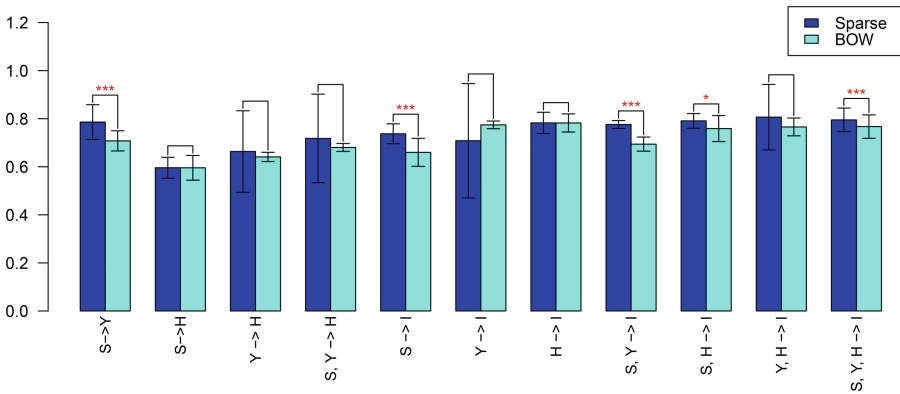
Fig. 2. Accuracy for predicting intent classes across different experimental settings. The X-axis represents the source to target events, where $S = Hurricane\ Sandy$, $Y = Super-typhoon\ Yolanda$, $H = Hurricane\ Harvey$, $I = Hurricane\ Irma$. Single *and* Triple Asterisks *show where the difference is significant at* 99% *and* 99.99%, *respectively*.

4 Experimental Setup

We employ two different experimental schemes for evaluating the performance of our proposed transfer learning method as follows:

– **Single past disaster as source**: In this case, we use only one past disaster dataset to learn the features for the source training set and a future disaster event dataset is used as a target test set.
– **Multiple past disasters as source**: In this case, we use more than one past disaster event datasets with different combinations as the source training set and a future disaster event dataset is used as a target test set.

Given the different possible selections of train/test set, we chose to consider only the case of predicting on the *future* disaster event given the source of *past* events on the timeline (*c.f.* Table 2). For evaluating the effectiveness of the proposed Sparse feature representation, we created a baseline of BoW representation for the features. The results are provided by repeating the experiments 20 times. A Wilcoxon's Signed-Rank test is used to find significant contrasts between the results of BOW and Sparse representation at 99% level and the significant cases are marked in the figures and tables by the *asterisks*.

5 Results and Discussion

This section discusses experimental results for the defined schemes of single and multi-event sources as well as the benefits of the Sparse Coding representation.

5.1 Performance Analysis

Figure 2 shows the results for both single and multi-source experiment settings. Overall, We note the following key observations from the results:

1. The accuracy was achieved close to 80%, which shows good predictability for the proposed model, given the complexity of detecting both explicit and implicit intent expressions from textual utterances.
2. We generally found the superior performance of the Sparse representation in contrast to the BoW representation across both single and multi-source experiments. In the setting of the source dataset as Yolanda and the target as Irma, the poor performance is likely due to the small size of the Yolanda dataset for training. We suspect this role of the small dataset in influencing the performance through the multi-source type experiments.
3. We note better performance in the experiments of leveraging multiple event datasets as the source in contrast to only the single event data source, which is likely contributed by the ability of Sparse representation to effectively capture intent cues across diverse disaster contexts.

We further observe the following points from Table 3:

Table 3. Details of the experimental setup and performance for F-score metric. K shows the number of atoms and λ holds for the regularization term in Sparse Coding.

Train	Test	Scheme	K	λ	F-score		
					Seeking	None	Offering
Sandy	Yolanda	Sparse	22	0.009	**0.744 ± 0.072***	**0.831 ± 0.092***	**0.682 ± 0.055***
		BOW			0.681 ± 0.037	0.758 ± 0.053	0.559 ± 0.035
Sandy	Harvey	Sparse	74	0.02	**0.529 ± 0.026***	0.719 ± 0.048	**0.187 ± 0.016**
		BOW			0.501 ± 0.027	**0.720 ± 0.060**	0.172 ± 0.017
Yolanda	Harvey	Sparse	37	0.0005	0.425 ± 0.107	**0.797 ± 0.235**	**0.174 ± 0.031***
		BOW			**0.515 ± 0.017***	0.769 ± 0.017	0.146 ± 0.013
Sandy	Irma	Sparse	17	0.007	**0.457 ± 0.040**	**0.847 ± 0.034***	**0.170 ± 0.015***
		BOW			0.443 ± 0.040	0.786 ± 0.056	0.146 ± 0.017
Yolanda	Irma	Sparse	12	0.0015	0.449 ± 0.154	0.837 ± 0.284	**0.188 ± 0.037**
		BOW			**0.535 ± 0.029**	**0.879 ± 0.010**	0.119 ± 0.022
Harvey	Irma	Sparse	55	0.0085	**0.605 ± 0.040**	**0.882 ± 0.034**	0.276 ± 0.025
		BOW			0.583 ± 0.026	0.879 ± 0.029	**0.316 ± 0.029***
Sandy, Yolanda	Harvey	Sparse	57	0.01	0.520 ± 0.126	0.768 ± 0.279	**0.172 ± 0.033***
		BOW			**0.546 ± 0.017**	**0.794 ± 0.014**	0.128 ± 0.012
Sandy, Yolanda	Irma	Sparse	78	0.0007	**0.571 ± 0.023***	**0.881 ± 0.012***	**0.159 ± 0.012***
		BOW			0.501 ± 0.032	0.815 ± 0.023	0.141 ± 0.016
Sandy, Harvey	Irma	Sparse	87	0.007	0.504 ± 0.026	**0.889 ± 0.022**	**0.303 ± 0.016***
		BOW			**0.527 ± 0.035***	0.866 ± 0.046	0.257 ± 0.023
Yolanda, Harvey	Irma	Sparse	90	0.0003	0.397 ± 0.085	**0.898 ± 0.133**	**0.297 ± 0.025**
		BOW			**0.585 ± 0.023***	0.870 ± 0.029	0.288 ± 0.025
Sandy, Yolanda, Harvey	Irma	Sparse	90	0.0015	**0.594 ± 0.041**	**0.890 ± 0.038***	0.273 ± 0.023
		BOW			0.568 ± 0.036	0.870 ± 0.040	**0.286 ± 0.019***

1. A general pattern for F-score indicates the better performance of Sparse representation than BoW in predicting important classes of help *offering* and *seeking*. The potential factor for the varied performance of Sparse Coding representation in Table 3 is the difference in the labeled sample size and class distribution of source datasets (as shown in Table 2).
2. We observe the low predictive power for the *offering* class, which is likely affected by the imbalanced class distribution and the lowest number of labeled instances for this class across each dataset.
3. We further note a direct relation between the size of source data and the optimal number of Atoms (K) that is discovered using Cross-Validation process as explained in Sect. 3.2 and shown in Fig. 1. Given a larger dataset may contain more information, it would likely require more bases for information representation.

5.2   Limitation and Future Work

We presented a novel approach of transfer learning for efficiently identifying help-seeking or offering intent in a future disaster event, in contrast to event-specific supervised learning methods requiring large labeled datasets for the future event. While we presented the experiments for disaster events of similar types (hurricane and typhoon), future work could investigate the performance of transfer learning of help intent across the types of disasters, such as earthquakes and hurricanes. We considered English language tweets due to the complexity of understanding intent from the short text and our natural next step is to leverage the proposed method for cross-language intent identification. Further, the proposed method provides a general framework for boosting any input text representation by adding a layer of Sparse Coding using the explained workflow in Fig. 1. For instance, a future study can enhance the learning performance while using more advanced input representations such as n-grams, word vectors, etc. for different types of learning tasks.

6   Conclusion

This paper presented a novel approach of transfer learning with Sparse Coding feature representation for the task of help intent identification on social media during disasters. We experimented with four disaster event datasets ordered over time and analyzed the performance of the proposed model for both single-event source and multi-events source schemes to predict in a target (future) event. Our results showed that using the Sparse Coding representation model in contrast to the baseline model of Bag-of-Words representation allows the efficient transfer of knowledge for intent behaviors from past disaster datasets. The application of the proposed approach can enhance real-time social media filtering tools during future disaster responses, when there would be insufficient event-specific labels to train a supervised learning model.

References

1. Abbasi, M.-A., Liu, H.: Measuring user credibility in social media. In: Greenberg, A.M., Kennedy, W.G., Bos, N.D. (eds.) SBP 2013. LNCS, vol. 7812, pp. 441–448. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37210-0_48
2. Ashkan, A., Clarke, C.L.A., Agichtein, E., Guo, Q.: Classifying and characterizing query intent. In: Boughanem, M., Berrut, C., Mothe, J., Soule-Dupuy, C. (eds.) ECIR 2009. LNCS, vol. 5478, pp. 578–586. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-00958-7_53

3. Castillo, C.: Big Crisis Data: Social Media in Disasters and Time-Critical Situations. Cambridge University Press, Cambridge (2016)
4. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006)
5. Faruqui, M., Dodge, J., Jauhar, S.K., Dyer, C., Hovy, E., Smith, N.A.: Retrofitting word vectors to semantic lexicons. In: Proceedings of NAACL (2015)
6. Faruqui, M., Tsvetkov, Y., Yogatama, D., Dyer, C., Smith, N.A.: Sparse overcomplete word vector representations. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (vol. 1: Long Papers), pp. 1491–1500 (2015)
7. Han, Y., Wu, F., Zhuang, Y., He, X.: Multi-label transfer learning with sparse representation. IEEE Trans. Circ. Syst. Video Technol. **20**(8), 1110–1121 (2010)
8. He, X., Lu, D., Margolin, D., Wang, M., Idrissi, S.E., Lin, Y.R.: The signals and noise: actionable information in improvised social media channels during a disaster. In: Proceedings of the 2017 ACM on Web Science Conference, pp. 33–42. ACM (2017)
9. Huang, Z., Pan, Z., Lei, B.: Transfer learning with deep convolutional neural network for SAR target classification with limited labeled data. Remote Sens. **9**(9), 907 (2017)
10. Hughes, A.L., St Denis, L.A., Palen, L., Anderson, K.M.: Online public communications by police & fire services during the 2012 hurricane sandy. In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, pp. 1505–1514. ACM (2014)
11. Imran, M., Castillo, C., Diaz, F., Vieweg, S.: Processing social media messages in mass emergency: a survey. ACM Comput. Surv. (CSUR) **47**(4), 67 (2015)
12. Imran, M., Elbassuoni, S., Castillo, C., Diaz, F., Meier, P.: Practical extraction of disaster-relevant information from social media. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 1021–1024. ACM (2013)
13. Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. In: Advances in Neural Information Processing Systems, pp. 801–808 (2007)
14. Lee, H., Raina, R., Teichman, A., Ng, A.Y.: Exponential family sparse coding with application to self-taught learning. In: IJCAI, vol. 9, pp. 1113–1119 (2009)
15. Maurer, A., Pontil, M., Romera-Paredes, B.: Sparse coding for multitask and transfer learning. In: Proceedings of the 30th International Conference on Machine Learning (ICML 2013), pp. 343–351 (2013)
16. Nazer, T.H., Morstatter, F., Dani, H., Liu, H.: Finding requests in social media for disaster relief. In: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 1410–1413. IEEE (2016)
17. 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 & Social Computing, pp. 994–1009. ACM (2015)
18. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2010)
19. Peng, F., Schuurmans, D.: Combining naive bayes and n-gram language models for text classification. In: Sebastiani, F. (ed.) ECIR 2003. LNCS, vol. 2633, pp. 335–350. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-36618-0_24
20. Plotnick, L., Hiltz, S.R.: Barriers to use of social media by emergency managers. J. Homel. Secur. Emerg. Manag. **13**(2), 247–277 (2016)

21. Purohit, H., Castillo, C., Diaz, F., Sheth, A., Meier, P.: Emergency-relief coordination on social media: automatically matching resource requests and offers. First Monday **19**(1) (2013)
22. Purohit, H., Dong, G., Shalin, V., Thirunarayan, K., Sheth, A.: Intent classification of short-text on social media. In: 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity), pp. 222–228. IEEE (2015)
23. Purohit, H., Hampton, A., Bhatt, S., Shalin, V.L., Sheth, A.P., Flach, J.M.: Identifying seekers and suppliers in social media communities to support crisis coordination. Comput. Support. Coop. Work (CSCW) **23**(4–6), 513–545 (2014)
24. 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 (2011)
25. Quattoni, A., Collins, M., Darrell, T.: Transfer learning for image classification with sparse prototype representations. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, pp. 1–8. IEEE (2008)
26. Reuter, C., Ludwig, T., Kaufhold, M.A., Spielhofer, T.: Emergency services' attitudes towards social media: a quantitative and qualitative survey across europe. Int. J. Hum. Comput. Stud. **95**, 96–111 (2016)
27. Reuter, C., Spielhofer, T.: Towards social resilience: a quantitative and qualitative survey on citizens' perception of social media in emergencies in Europe. Technol. Forecast. Soc. Change **121**, 168–180 (2016)
28. Rodriguez, F., Sapiro, G.: Sparse representations for image classification: learning discriminative and reconstructive non-parametric dictionaries. Technical report, Minnesota Univ Minneapolis (2008)
29. Singh, O.P., Haris, B., Sinha, R.: Language identification using sparse representation: a comparison between GMM supervector and i-vector based approaches. In: 2013 Annual IEEE India Conference (INDICON), pp. 1–4. IEEE (2013)
30. Sparck Jones, K.: A statistical interpretation of term specificity and its application in retrieval. J. Doc. **28**(1), 11–21 (1972)
31. Starbird, K., Palen, L.: Pass it on?: retweeting in mass emergency. In: Proceedings of the 7th International ISCRAM Conference-Seattle, vol. 1. Citeseer (2010)
32. Stephenson, J., Vaganay, M., Coon, D., Cameron, R., Hewitt, N.: The role of Facebook and Twitter as organisational communication platforms in relation to flood events in Northern Ireland. J. Flood Risk Manag. (2017)
33. Sutton, J.N., Spiro, E.S., Johnson, B., Fitzhugh, S.M., Greczek, M., Butts, C.T.: Connected communications: network structures of official communications in a technological disaster. In: Proceedings of the 9th International ISCRAM Conference (2012)
34. Varga, I., Sano, M., Torisawa, K., Hashimoto, C., Ohtake, K., Kawai, T., Oh, J.H., De Saeger, S.: Aid is out there: looking for help from tweets during a large scale disaster. In: ACL, vol. 1, pp. 1619–1629 (2013)
35. Vieweg, S.E.: Situational awareness in mass emergency: a behavioral and linguistic analysis of microblogged communications. Ph.D. thesis, University of Colorado at Boulder (2012)