Newswire versus Social Media for Disaster Response and Recovery

Rakesh Verma
Department of Computer Science
University of Houston
Houston, United States
rverma@uh.edu

Samaneh Karimi
School of Electrical and Computer Engineering
University of Tehran
University of Houston, Houston, United States
samanekarimi@ut.ac.ir
skarimi2@uh.edu

Daniel Lee
Department of Computer Science
University of Houston
Houston, United States
dljr0122@cs.uh.edu

Omprakash Gnawali
Department of Computer Science
University of Houston
Houston, United States
gnawali@cs.uh.edu

Azadeh Shakery
School of Electrical and Computer Engineering
University of Tehran
Institute for Research
in Fundamental Sciences (IPM), Tehran, Iran
shakery@ut.ac.ir

***Abstract*—In a disaster situation, first responders need to quickly acquire situational awareness and prioritize response based on the need, resources available and impact. Can they do this based on digital media such as Twitter alone, or newswire alone, or some combination of the two? We examine this question in the context of the 2015 Nepal earthquakes. We propose a method to link tweets and newswire, so that we can compare their key characteristics: timeliness, whether tweets appear earlier than their corresponding newswires, and content. Whenever possible, we present both quantitative and qualitative evaluations. One of our main findings is that tweets and newswire articles provide complementary perspectives that form a holistic view of the disaster situation. Another finding is that during the Nepal earthquake, significant information of earthquake-related news appeared on Twitter *before* newswire. Our results also show that 9.08% of the tweets published before their matched newswire contain actionable information regarding the Nepal earthquake.**

***Index Terms*—Nepal earthquake, newsline, Twitter, critical infrastructure resilience**

I. Introduction

When a disaster strikes, responders and relief agencies need to rapidly assess the damages to lives and infrastructures, and get a grip on the situation. Phone service and electricity supply may be disrupted in various parts of the affected region. Thus, there may not be direct sources of information available, e.g., calling (or messaging) the first responders in the affected region may not be possible. How can important and reliable information be acquired quickly in such a fast-moving and chaotic situation? For this purpose, they may turn to indirect sources: the social networks, such as Twitter, or the newswire services.

Twitter has become a *de facto* standard domain for event detection [1], because of its real-time nature. Researchers have given some evidence to show that Twitter users break news before newswire. However, few studies have examined the content and timeliness of the two different sources, especially in the context of a major disaster.

In this paper, we examine the content and timeliness of the two different sources in the context of the 2015 Nepal earthquakes. These considerations lead to the following research questions (RQ):

1) Does Twitter report news faster than traditional newswire, especially in the context of a rapidly changing situation such as a major disaster? Note that this question is more interesting when some content is common to both tweet and the article? (RQ1)
2) What type of information is reported earlier by Twitter, especially in the context of a rapidly changing situation such as a major disaster? Again, this question is more interesting when the tweet and the article have something in common. (RQ2)
3) How effective are some of the methods proposed for linking tweets with newswire? One motivation is to compare and analyze their characteristics especially in the context of major disasters. Another motivation stems from RQ1 and RQ2. Without linking the tweets and the news articles, we would be simply reporting that a tweet appeared before any news article, which is not too interesting. (RQ3)

A preliminary version of our work appears in [2]. The rest of this paper is organized as follows. In Section II, we describe the datasets. Section III describes the proposed method for linking tweets with newswires and its evaluation. Section IV compares Twitter and newswire to address RQ1 and RQ2. We then discuss the linking method's effectiveness (RQ3) in Section V. In Section VI, we review the relevant related work and Section VII concludes.

II. DATASETS

The two datasets, a tweet and a newswire dataset, we collected for the research questions are described below.

A. Twitter Dataset & Characteristics

There exist some tweet datasets on the Nepal earthquake, e.g., [3]. However, since the main goal in this paper is to use the linkage between tweets and newswire content for disaster response and recovery, we need two *contemporaneous* datasets of newswire and tweets, relevant to the Nepal earthquake. Therefore, we collected a set of tweets about the Nepal earthquake using "Nepal earthquake" as the search query and annotated them. We call this dataset NE-Tweets. We also collected newswire data as explained in Section II-B.
Preprocessing: All tweets are preprocessed by removing non-ascii characters, punctuation, hashes, stopwords and URLs.

NE-Tweets consists of 336,140 tweets written from April 24, 2015 to June 25, 2015. Exploratory and descriptive statistics of NE-Tweets dataset are as follows.

1) Most Frequent Words: As shown in Table I, the top five most frequent words in NE-Tweets are *nepal*, *earthquake*, *help*, *relief* and *victims*. The vocabulary size of NE-Tweets, i.e., number of unique words, is 91,752. The careful reader might think that the words *nepal* and *earthquake* should be in every tweet since that is how the dataset was created, but some tweets had #earth#quake, which after removing hashes became "earth quake," hence the problem. This issue was discovered by examining the tweets that did not have any disaster-related keyword.

TABLE I
MOST FREQUENT WORDS IN NE-TWEETS

Word	No. of tweets	Percentage
nepal	329,836	98.1%
earthquake	312,721	93.0%
help	31,012	9.2%
relief	26,062	7.8%
victims	24,823	7.4%

2) Time versus Number of Tweets: The number of tweets in NE-Tweets varies over time as shown in Figure 1. The higher number of tweets during the first two periods, i.e. from April 24 to May 13, is due to the occurrence of the earthquake on April 25, its aftershocks and all related issues that were raised consequently. Notice the sharp drop in Twitter activity in NE-Tweets 20 days after the earthquake.

3) Twitter Users Activity: The tweets in NE-Tweets are generated by 160,053 unique users, i.e., on the average slightly over two tweets per user. The top four users ranked based on the number of tweets authored in NE-Tweets are shown in Table II. According to their profiles, the top ranked users are those who are active in news area in related domains such as Nepal related news, natural disasters or Muslim world news.

4) The Descriptive Statistics of the Twitter Dataset: Table III shows some of the descriptive statistics of the collected tweet dataset including the fraction of tweets with URLs, mentions, hashtags and disaster keywords. To compute the fraction of tweets that contain a disaster keyword, a list of keywords about weather, disaster and emergency[1] is employed. We see that 99.4% of tweets contain at least one disaster keyword.

TABLE II
THE TOP 4 MOST ACTIVE USERS IN NE-TWEETS

Username	Description	No. of Tweets	%-age of Tweets
dlnepalnews	Link you with news in Nepal.	1,601	0.48%
gcmcEarthquake	We tweet about Crisis, Disaster and Emergency Management related to Earthquakes.	1,078	0.32%
crowdtrendies	The latest uber great campaigns from all your favourite crowd funding websites.	752	0.22%
wilayah_news	Breaking news and information from the Muslim world.	673	0.20%
Totals for 4 usernames		4,104	1.22%

TABLE III
DESCRIPTIVE STATISTICS OF NE-TWEETS

Percentage of tweets with mentions	17.4%
Percentage of tweets with URLs	77.4%
Percentage of tweets with at least one disaster keyword	99.4%
Percentage of tweets with hashtags	32.1%

B. Newswire Dataset & Characteristics

News articles were collected from five important Nepali news sources: "Kantipur," "Kathmandu Post," "Nayaptika," "Nepali Times" and "The Rising Nepal." In total more than 700 articles were obtained in both English and Nepali, all of them published between April 2, 2015 and November 5, 2015. The English subset ranges from April 28, 2015 to August 19, 2015. We call this the ENE-news dataset.

1) Some Challenges of ENE-News: Because the articles were printed in Nepal, even ENE-News had encoding that was incompatible with standard English characters. So the articles in ENE-News were filtered further to include only those that can be decoded to ASCII characters.

After this, another round of filtering was done to ignore articles which contained less than 1000 characters (approx. 100 words). And the final dataset used for annotation required a minimum of 10 sentences. We call this the ENE-News-final dataset.

TABLE IV
DETAILS OF FILES REMAINING AFTER FILTERING THE NEWSWIRE DATASET.

Filter	Articles Remaining	Aux. Info
English-Only	799	5 were empty files
1000 character minimum	517	414 avg. words per article
10 sentence minimum	349	519 avg. words per article

2) Preprocessing: For consistency, all articles in ENE-News-final were preprocessed as follows:

- Newswire content was parsed into sentences, and then each sentence into word tokens.

[1] https://gist.github.com/jm3/2815378

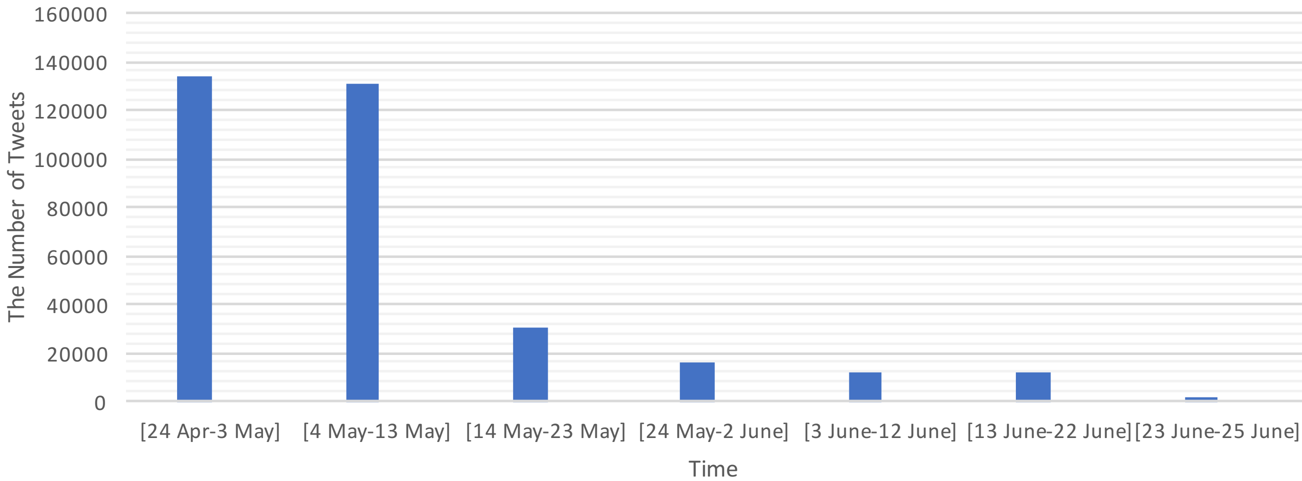

Fig. 1. The number of tweets written in NE-Tweets over 10 day periods from 24 Apr to 25 June.

- Stopwords (words with low information value) were removed.
- Stemming was done to allow words like “work,” “worked” and “working” to be considered as the same.
- During any comparison, the word tokens were always lowercased.

The preprocessing was done using the open-source package Natural Language Toolkit (NLTK) [4].

C. Tweet-Newswire Annotation

To evaluate the performance of the proposed tweet-newswire linking method, a set of 310 tweet-newswire pairs were selected to be annotated by a group of 10 researchers, such that each pair was annotated by two annotators. This set of 310 pairs is the result of pairing a set of 31 newswires and their top 10 most similar tweets based on their TFIDF similarity score. Therefore, the set of pairs may contain duplicate tweets. The annotators were asked to annotate each tweet-newswire pair according to the following instructions (along with an example of each label type).

- If the tweet is relevant to the newswire, i.e. it is about a specific subject that is also mentioned in the newswire, it should be labeled as relevant (label = 2).
- If the tweet is generally relevant to the topic of the newswire, it should be labeled as partially relevant (label = 1).
- If the tweet doesn’t have a meaningful content or is totally irrelevant to the newswire, it should be labeled as not relevant (label = 0).

III. Linking Tweets with Newswire

To link the tweets in NE-Tweets with articles in ENE-News-final, a supervised machine learning approach is proposed that explores the space of tweet-newswire article pairs. In this approach, each pair of tweet and newswire is represented by a set of features, then a classification model is learned using a training set with matching labels. Finally, the trained model is applied on the tweet-newswire pairs of the test set to find the matched ones. The features for each tweet-newswire pair are as follows.

char Ngram similarity score: This feature measures the similarity between the char Ngrams of the tweet and the newswire by counting the number of the matched char Ngrams normalized by the total number of possible matched character Ngrams. The rationale for this feature is to detect the similarities between tweet’s words, specially hashtags, written in camel case style and newswire words. In this paper, this feature is computed for $N = 2$ and $N = 3$. Furthermore, the same feature is computed for the expanded versions of both the tweet and newswire datasets. For expansion, we add all WordNet synsets [5] of each word found in the text. Thus, four different features are computed for each tweet-newswire pair including char2gramSim, char3gramSim, exp_char2gramSim and exp_char3gramSim.

Temporal distance: The difference between the publish date of the newswire and the tweet in days, after converting their different time zones into the same one.

TFIDF score: The TFIDF similarity between the tweet and newswire content. The TFIDF similarity scores are calculated using the Lemur project,[2] where tweets are treated as queries to a corpus of newswires. The use of this feature in our task is similar to the idea of using TFIDF similarity scores for linking the citation text in citing papers to cited papers in [6].

Hashtag similarity: It is the number of hashtags matched with any term in the newswire, normalized by the total number of hashtags used in the tweet.

Since tweet-article pairs are used as instances and the number of matched pairs is very few compared to all possible pairs of tweets and newswire in the training set, the training

[2] https://www.lemurproject.org

data is imbalanced. Therefore, a random undersampling is performed by randomly picking samples from the majority class without replacement to make the training data balanced before learning. Support Vector Machine (SVM) with a Radial Basis Function kernel is used as the classification method.

A. Experimental Results

1) Training and Test Sets: As mentioned in section II-A, tweets related to the Nepal earthquake were collected. Then, the list of keywords about weather, disaster and emergency, mentioned in II-A4, is used to filter the tweets that contain at least one of the keywords by exact matching. The final set of tweets is employed as the test set. We used another dataset of tweets and newswires with their matching labels [7] as the training set. The number of tweets in the training set is 34,888 and the number of newswires in the training set is 12,704. We created the feature vectors for all pairs of tweets and their top (up to) 100 retrieved newswires, based on their TFIDF similarity score, for both training and test sets. In summary, the total number of instances, i.e. tweet-newswire pairs, in the training data is 759,971 and in the test data is 528,402.

2) Experiments: We used the Scikit-Learn and imblearn Python libraries for the implementation of the classification and undersampling methods respectively. We randomly selected one fifth of the training set and used it as the validation set for tuning γ and C, the parameters of the SVM method. Higher γ values leads to shorter radius of influence of support vectors. C acts as regularization parameter in SVM that controls the width of the decision margin. To tune the parameters, we performed 5-fold cross validation on the validations set using the GridSearchCV module of scikit-learn library.

For each newswire, we ranked the tweets based on their class membership probabilities, namely the probabilities of being classified as relevant, estimated by SVM. We selected the top 10 tweets for each newswire and annotated the resulting pairs as relevant, partially relevant or irrelevant for a subset of 31 newswires from the test set. To aggregate the two labels assigned by two annotators for each pair, we considered the ceiling of their arithmetic mean as the final annotation for that pair. In other words, we look at the sum of the two labels; if it is 0, the final annotation is 0 (i.e. irrelevant pair), if 1 or 2, the final annotation is 1 (i.e. partially relevant pair) and if 3 or 4, the final annotation is 2 (i.e. relevant pair). After obtaining the aggregated annotations for each pair, the precision is calculated. To compute the precision, we considered the partially relevant examples as true positive with a weight of 0.5. The precision and aggregated annotation results on the set of 310 tweet-newswire pairs are shown in Table V.

TABLE V
THE PRECISION AND ANNOTATION RESULTS ON THE SET OF 310 PAIRS.

The number of relevant pairs	The number of partially relevant pairs	Precision
37	218	0.47

We computed the actual agreement between the two annotators by considering a weight of 0.5 for which the annotations difference equals to one. The computed score for our annotations is 0.59 which means that there is a moderate agreement between the annotators.

B. Some Challenges of the Linking Task

Some of the challenges of the tweet-newswire linking task are:

- The lack of published time information of newswires.
- The lack of geographic location information of tweets.
- Different time zones in tweets' time information.

IV. Comparing Twitter and Newswire Information

A. News Reporting Speed

One of the aspects of comparing Twitter and traditional newswire is their speed in reporting news, specially during a disaster. To this aim and to investigate the first research question (RQ1), we used the set of tweet-newswire pairs classified as matched by the proposed method and also annotated as relevant or partially relevant pairs by annotators. For these pairs, we calculated the temporal distance between the tweet and its corresponding newswire by subtracting the tweet publish date from newswire publish date. We computed the temporal distances for all matched pairs annotated as relevant and partially relevant separately. Figures 2 and 3 show the histograms of the temporal distances of the matched pairs annotated as relevant and partially relevant by bin. The size of the bins is 5.

As shown in Figures 2 and 3, in most cases the temporal distances are positive which means in most matched pairs of tweet-newswire, the newswire publish date is older than its matched tweet date. Furthermore, the percentage of tweets that appeared before their matching newswires are 91.8% and 91.7% in relevant and partially relevant pairs, respectively. This implies that news is reported by tweets faster than newswires in both relevant and partially relevant pairs.

B. Clustering on Tweets - Discovering Events

To investigate the second research question (RQ2), we used the same set of tweets that was used in Section IV-A, then we removed the small number of tweets that were published after their matched newswires and employed the resulting tweet set for clustering. In other words, we obtained a set of tweet-newswire pairs that were classified as matched by our tweet-newswire linking method, annotated as relevant or partially relevant by annotators and were published before their matched newswire to find what type of information is reported *earlier* by Twitter by visualizing the clusters found.

For clustering tweets, we used a short text clustering method called GSDMM which is a collapsed Gibbs Sampling algorithm for the Dirichlet Multinomial Mixture model [8]. In this experiment, the values of the GSDMM parameters are $k = 15$, $\alpha = 0.1$, $\beta = 0.1$, $t = 50$ where k is the number of clusters, α and β are Dirichlet priors and t is the number of iterations that the clustering algorithm repeats until convergence. Since,

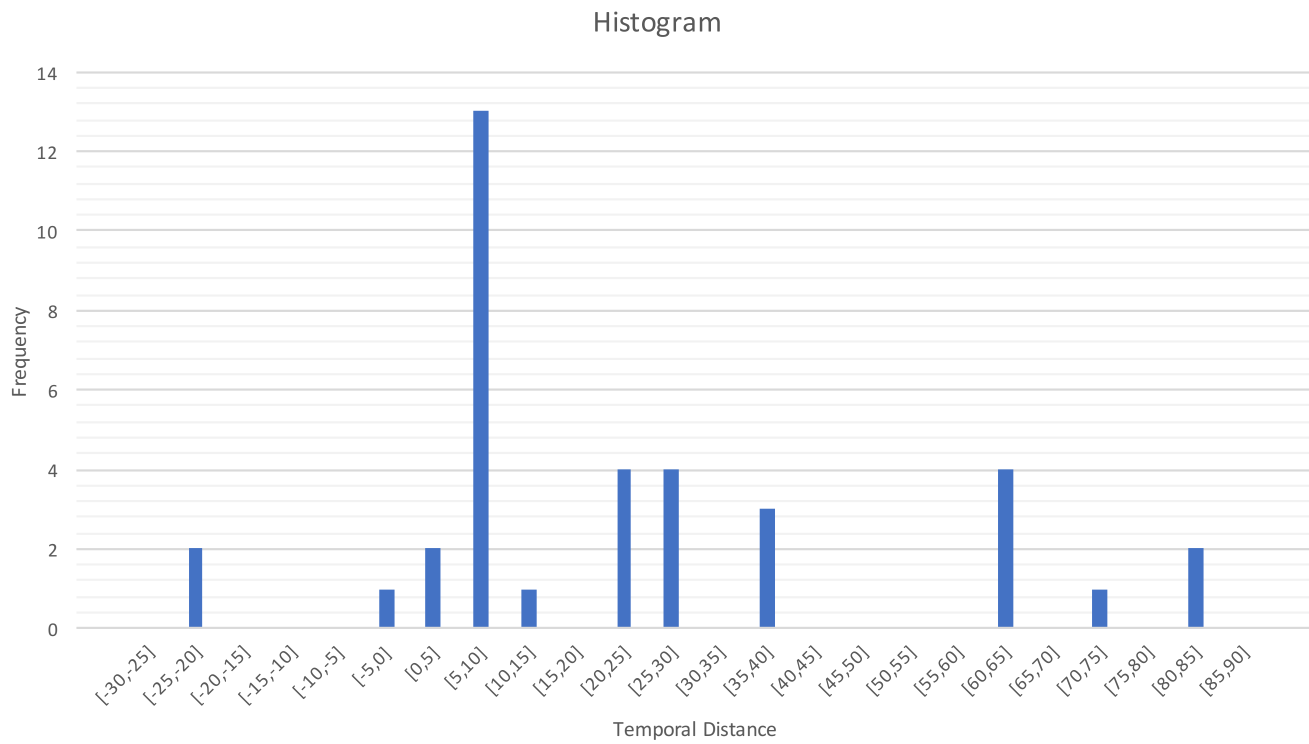

Fig. 2. The histogram of the temporal distances of the matched pairs annotated as relevant.

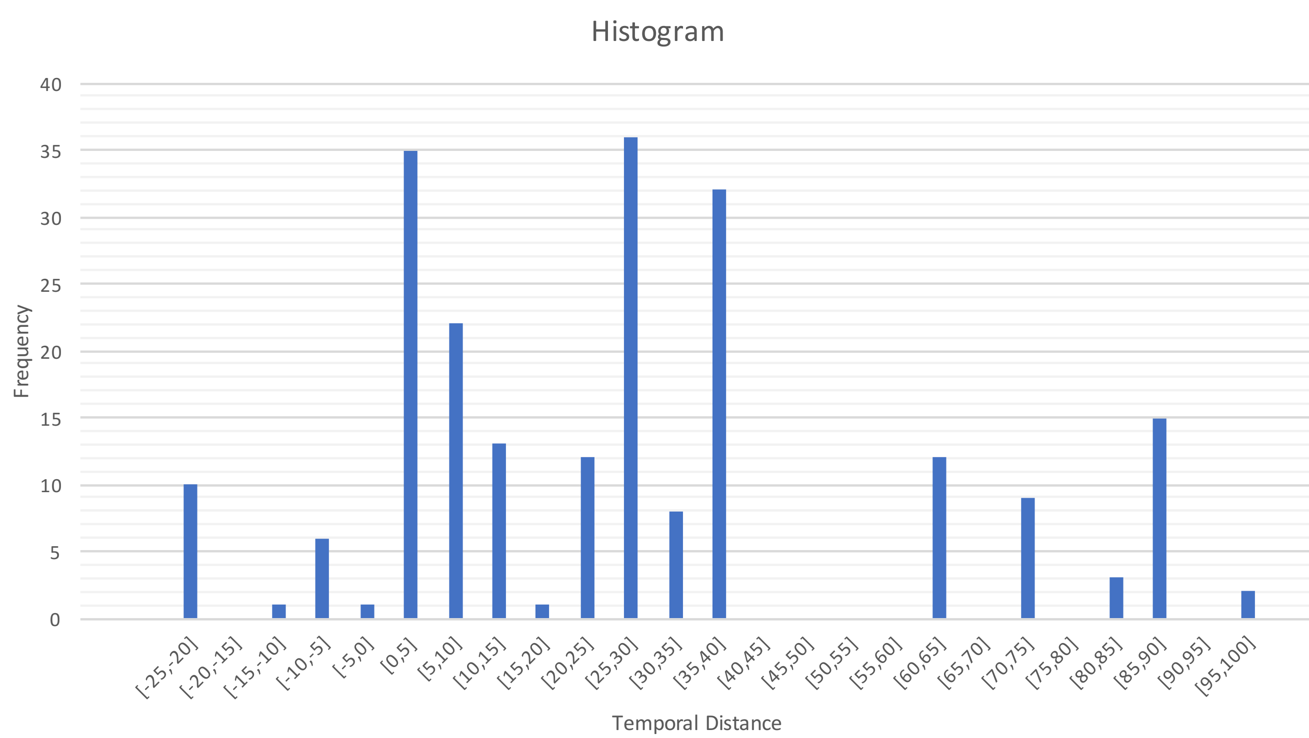

Fig. 3. The histogram of the temporal distances of the matched pairs annotated as partially relevant.

the ground truth clusters of our tweets are not available, we manually checked the quality of clusters using multiple k values and selected $k = 15$ as the best value. We also used the default values of α, β in [8]. The total number of tweets used as input in this experiment is 172. Table VI contains some statistics about the clustering result. In this experiment, tweets are preprocessed by removing the non-ascii characters, punctuation, stopwords and URLs.

TABLE VI
SOME STATISTICS ABOUT THE CLUSTERING RESULTS ON TWEETS.

Min cluster size	Max cluster size	Average Cluster size	Mode of the cluster sizes
8	18	11.47	9

Some of the clusters obtained from applying GSDMM on the set of tweets mentioned above are shown in Figure 4. The word cloud representation is used to show the clusters. As Figure 4 shows, each of these clusters correspond to a separate topic relevant to the Nepal earthquake. The topics corresponding to each of these clusters include a four month old baby being pulled from the rubble (the top left cluster), a man being pulled from the rubble 82 hours after the Nepal earthquake (the top right cluster), India's effort to help the earthquake survivors (the bottom left cluster) and the health water crisis in Nepal after the earthquake (the bottom right cluster).

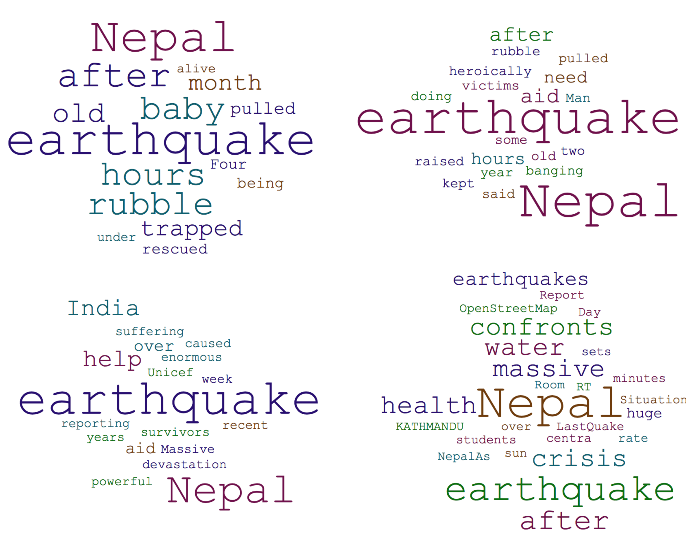

Fig. 4. The events obtained by GSDMM method using the tweets that appeared before their corresponding newswires.

C. Content Comparison

In this section, the word cloud representation is used to compare the textual content of newswires and the tweets content for relevant and partially relevant annotated pairs as shown in Figures 5, 6 and 7.

One of the differences of newswire and tweet content, shown in their word clouds is the different choice of words in the two channels. As Figure 5 show objective words such as 'government,' 'people' and 'houses' have high weights in newswires, while as shown in Figures 6 and 7, more opinionated and subjective words such as 'devastating,' 'massive' and 'suffering' are used in tweets. Another observation is that the newspaper articles have a lot of locations, e.g., 'Kathmandu,' 'Sindhupalchok,' 'Tamang,' and 'Barpak,' but the tweets focus more on the human angle, e.g., 'parents,' 'baby,' 'man,' 'month-old,' 'four-month-old,' 'year-old,' 'trapped,' 'pulled,' and 'rubble.'

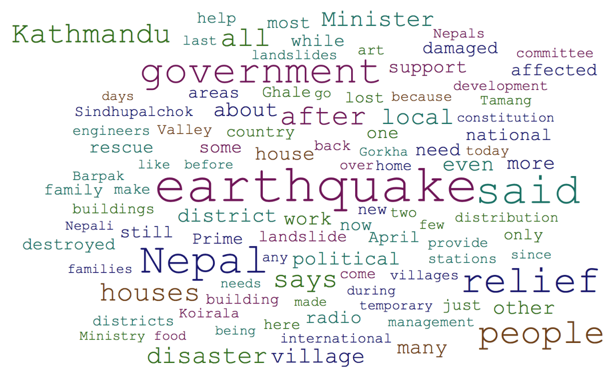

Fig. 5. The word cloud representation of the newswires' content.

Fig. 6. The word cloud representation of the tweets' content that were annotated as relevant.

D. The Extra Information in Tweets vs News

In the next experiment, we investigate the second research question (RQ2) by finding what new information is provided in tweets that is not available in newswires about the Nepal earthquake and vice versa, i.e. what new information is provided in newswires that is not available in tweets. In this experiment, we used the same 172 tweets and newswires used in Section IV-B. To this aim, we computed the difference of the aggregate of all tweets' content, denoted by T_{172}, from the aggregate of the matching newswires' content (NAC), i.e. $NAC - T_{172}$, and $T_{172} - NAC$, and represented the resulting set of words by their word clouds. Precisely speaking, to compute $NAC - T_{172}$, we find the common words between

Fig. 7. The word cloud representation of the tweets' content that were annotated as partially relevant.

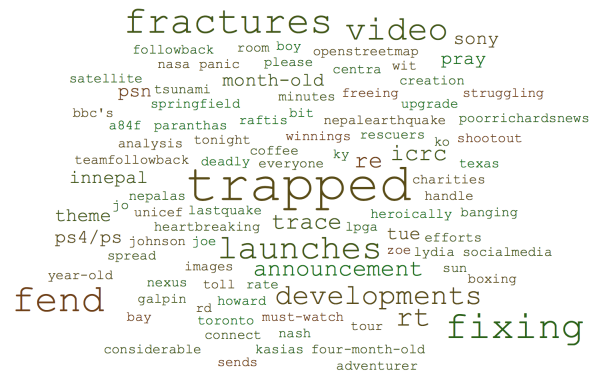

Fig. 9. Word cloud representation of $T_{172} - NAC$.

NAC and T_{172} and remove all of its occurrences from NAC. We do the same for computing $NAC - T_{172}$.

Comparing the two word clouds in Figures 8 and 9 shows that newswires mainly contain information about the Nepal earthquake from the agencies viewpoint. The use of terms such as 'government', 'reconstruction' and 'political' shows the formal and objective language employed by news agencies to describe the situation. In contrast, tweets content is more about the human angle, the content type that newswires lack. These word clouds reinforce that the two sources of information can complement each other to provide a comprehensive picture from two different angles about a major disaster like the Nepal earthquake. The immediacy of tweets can be used by first responders for search and rescue operations.

Fig. 8. Word cloud representation of $NAC - T_{172}$.

E. Analyzing Tweets' Information Types

In this experiment, the second research question (RQ2) is studied from another viewpoint by analyzing the types of information posted to Twitter before newswires during major disasters. To this aim, we randomly selected 7 tweets from each of the 15 clusters used in IV-B. Hence, we obtained a random sample of 105 tweets that are matched with newswires by our tweet-newswire linking method, annotated as relevant or partially relevant by annotators and were published before their matched newswire. Then, each of the tweets were annotated based on the information type they contain according to the major information types defined in [9]. In [9], 25 high-level types are defined to categorize the information shared on Twitter during emergencies. We employed four major categories of information types introduced in [9] including "Request", "CallToAction", "Report" and "Other." "Request" denotes that the tweet contains a request for a service, good, rescue or information. Tweets that include a request to the public for actions such as donating or volunteering to help should be annotated as "CallToAction". "Report" label should be assigned to tweets that only report some thing about the event. Other tweets that contain users' sentiments or discussions regarding the event or provide coverage of the event should be labeled as "Other." Each tweet is assessed by two annotators and we obtained moderate (0.417) Cohen's kappa agreement score among the annotators. Figure 10 shows the percentages of tweets labeled with each of the four information types in the subset of the sampled tweets that were annotated the same by both annotators.

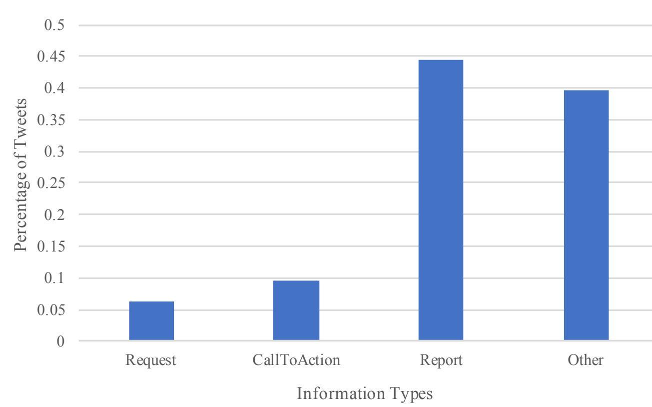

Fig. 10. Percentages of tweets, annotated the same by both annotators, by information type

As Figure 10 shows, the main information types reported in Twitter are "Report" (44.44%) and "Other" (39.68%) and the percentages of "Request" and "CallToAction" labels are 6.34%

and 9.52%, respectively. We can analyze the actionability of tweets for emergency situations based on their information types. "What is actionable to one person may not be the same as what is actionable to another" [10]. Therefore, for our analysis, we consider the "Request" and "CallToAction" information types as actionable. Because we assume that these two types of information can be potentially useful from the perspective of the first responder. According to our definition, 15.87% of the studied tweets contain actionable information.

In the next experiment, we used the information types as ground truth classes and compared with the 15 clusters used in IV-B. To this aim, we calculated purity [11] and V-Measure [12] and only considered the tweets with the same annotations from both annotators. The number of tweets with the same annotations were 63. The purity is 0.68 and the V-Measure is 0.23. These results show that each cluster that corresponds to a sub-topic relevant to the disaster, contains different information types. This finding is consistent with our expectations, since the clustering method was agnostic, i.e., did not use any specific features for clustering according to these categories. For example, the cluster which is about water crisis in Nepal after the earthquake may include tweets reporting the subject by an official (i.e. "Report" information type). Some other tweets of this cluster may contain victim's requests for clean water (i.e. "Request") or people's sentiments regarding this problem (i.e. "Other"). Therefore, the tweets' information types in each cluster can be different. A summary of all datasets used for each of the experiments are provided in table VII.

TABLE VII
THE DATASETS EMPLOYED FOR EACH EXPERIMENT.

ID	Experiment	Dataset
d1	Tweet-News linking (RQ3)	310 tweet-news pairs, the result of pairing a set of 31 newswires and their top 10 most similar tweets based on their TFIDF score.
d2	Comparing news reporting speed in Twitter and newswire (RQ1)	The set of tweet-news pairs classified as matched by the proposed method and also annotated as relevant or partially relevant pairs by annotators.
d3	Clustering on tweets (RQ2)	A subset of d2 whose tweets are all published before their matched newswires.
d4	The extra information in tweets vs news	Same as d3.
d5	Analyzing tweets information types (RQ2)	Randomly selected 7 tweets from each of the 15 clusters used in d3.

V. Discussion: Effectiveness of Tweet-Newswire Linking Model

We carefully analyzed the method proposed for linking tweets to newswire. With a precision of 0.47, the tweet-newswire linking is performing good enough for a deeper analysis of the content in the tweets. The collected data clearly shows that the tweets appear before the news articles. This fact in conjunction with the unique content reveals that tweets provide usable information that can be effectively employed for managing disaster situations.

Moreover, according to our results, the percentage of actionable tweets published before their matched newswire is considerable and can play an important role in serving emergency responders' information needs.

VI. Related Work

A. *Twitter for Actionable Information Amidst Disasters*

Twitter for emergency applications has been studied by several researchers, e.g., [2], [13]–[19]. In [13], researchers concluded that Twitter was not yet ready for first responders. However, it was helpful for civilians. These were the early days of Twitter, as we find from [14] that individuals immediately posted specific information helpful to "early recognition and characterization of emergency events" in the case of the Boston marathon bombing. In [15], researchers found that tangible, useful information was found in the early period before storm system Sandy and it got buried in emotional tweets as the storm actually hit. In [19], the authors first extract the tweets containing situational information using a set of vocabulary-independent features and an SVM classifier, then summarize the situational tweets using an Integer Linear Programming (ILP) framework. However, we think more studies are needed on this issue, since the tweets collected were rather small, approximately 27,000, using just the hashtag #sandy. A bilingual analysis of tweets obtained over 84 days overlapping the Tohoku earthquake showed, among other results, the correlation between Twitter data and earthquake events [16]. A survey of this literature can be found in [17]. Moreover, the trends and developments of 1339 papers published on the use of social media for crisis response and management is analyzed using automatic tools in [18].

Several papers have examined Twitter data in disasters [20], [21] and specifically in the Nepal earthquake context [22]–[26]. Relevance of tweets was examined by [22], [23]. However, note that our problem is different, viz., whether a tweet is relevant in the context of a given news article. The other papers examined different aspects such as public concerns and perceptions of disaster recovery efforts [24], [26] and public reaction to social media project of the police [25].

Researchers have examined the question of whether Twitter can replace newswire for breaking news [1]. They studied a period of 77 days in 2011 during which 27 events occurred. The biggest disasters in this event-set are: an airplane crash with 43 deaths, and a magnitude 5.8 earthquake in Virginia that caused infrastructural damage.[3] They collected a large dataset of tweets and newswires, but then eliminated a large collection of tweets based on clustering.

Thus, some researchers have focused on comparing the two sources of information, e.g., [1], some others utilize the joint information in them to improve the performance of news related tasks, and some papers try to discover the linkage between tweets and newswires [27]–[32].

[3]None of these disasters, bad as they are, rise to the level of the Nepal earthquake(s) of 2015 in which almost 10,000 lives were lost.

B. Tweet-Newswire Linking

In this section, the previous works on these areas are reviewed.

The tweet-newswire linking method proposed by Guo et. al, [33] with the aim of enriching short text data in social networks with a graph based latent variable model. They extract text-to-text relations using hashtags in tweets and named entities in newswires along with their temporal similarity.

In [34], a framework for connecting newswires to Twitter conversations is proposed using local cosine similarity, global cosine similarity, local frequency of the hashtag and global frequency of the hashtag as the classification features extracted for each article-hashtag pair. The task of linking tweets with related newswires is studied in another paper to construct user profiles [35]. The authors proposed two sets of strategies to find relevant newswires to each tweet in this paper. In addition to URL-based strategies, which is similar to the idea used in [36], they also proposed several content-based strategies that include computing the similarity between hashtag-based, entity-based and bag-of-word-based representations of tweets and newswires to discover the relation between them. In addition to user modeling, the tweet-newswire linking task has been employed in document summarization [37], sentiment analysis [38] and event extraction [39].

VII. Conclusion

In this paper, we studied, compared and analyzed newswire and Twitter from different viewpoints in the context of the 2015 Nepal earthquakes. In this regard, we collected and annotated two datasets: A tweet dataset that contains 336,140 tweets related to the Nepal earthquake written from April 24, 2015 to June 25, 2015 and a newswire dataset containing 700 newswires relevant to the Nepal earthquakes and dated in the year 2015. We presented descriptive statistics of the collected tweets from different viewpoints: the most frequent words, the top most active users, the changes in the number of tweets over time and the use of hashtags, URLs and mentions in the collected tweets.

Furthermore, using the tweet-newswire pairs classified as matched and annotated as relevant, we compared the speed of Twitter and newswire in news reporting (RQ1). We found that during the Nepal earthquake, most of the human news and earthquake related crises appear in Twitter *before* newswire. Another finding is that during a major disaster, Twitter contains more opinionated and subjective content in comparison with the newswire's content. We also show that Twitter data holds data that is complementary to the content of relevant newswires. For consumers of information like first responders, it is paramount that all available information about a natural disaster can be quickly processed. Our results also show that 9.08% of the tweets published before their matched newswire contain actionable information for the first responders (RQ2). Analyzing the main characteristics of actionable tweets and designing methods for detecting them automatically can better serve the first responders needs. We proposed a tweet-newswire linking method to find the matched tweets to each newswire and evaluated its performance using our annotated datasets (RQ3). It gave a decent precision of 0.47 on the annotated subset.

Acknowledgments

Research supported in part by NSF grants CNS 1319212, DGE 1433817, and DUE 1356705.

References

[1] S. Petrovic, M. Osborne, R. McCreadie, C. Macdonald, I. Ounis, and L. Shrimpton, "Can Twitter replace newswire for breaking news?" in *Seventh international AAAI conference on weblogs and social media*, 2013.

[2] R. Verma, S. Karimi, D. Lee, O. Gnawali, and A. Shakery, "Newswire versus social media for disaster response and recovery," *arXiv preprint arXiv:1906.10607*, 2019.

[3] F. Alam, S. Joty, and M. Imran, "Domain adaptation with adversarial training and graph embeddings," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 1077–1087. [Online]. Available: https://www.aclweb.org/anthology/P18-1099

[4] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python*. O'Reilly Media, Inc., 2009.

[5] G. A. Miller, "Wordnet: A lexical database for English," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[6] S. Karimi, L. Moraes, A. Das, A. Shakery, and R. Verma, "Citance-based retrieval and summarization using IR and machine learning," *Scientometrics*, vol. 116, no. 2, pp. 1331–1366, Aug 2018.

[7] W. Guo, H. Li, H. Ji, and M. Diab, "Linking tweets to news: A framework to enrich short text data in social media," in *Proceedings of the 51th Annual Meeting of the Association for Computational Linguistics*, 2013.

[8] J. Yin and J. Wang, "A dirichlet multinomial mixture model-based approach for short text clustering," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2014, pp. 233–242.

[9] R. Mccreadie, C. Buntain, and I. Soboroff, "Trec incident streams: Finding actionable information on social media," in *Proceedings of ISCRAM*, 2019.

[10] H. Zade, K. Shah, V. Rangarajan, P. Kshirsagar, M. Imran, and K. Starbird, "From situational awareness to actionability: Towards improving the utility of social media data for crisis response," *Proceedings of the ACM on Human-Computer Interaction - CSCW*, vol. 2, no. CSCW, pp. 195:1–195:18, Nov. 2018. [Online]. Available: http://doi.acm.org/10.1145/3274464

[11] C. Manning, P. Raghavan, and H. Schütze, "Introduction to information retrieval," *Natural Language Engineering*, vol. 16, no. 1, pp. 100–103, 2010.

[12] A. Rosenberg and J. Hirschberg, "V-measure: A conditional entropy-based external cluster evaluation measure," in *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*, 2007.

[13] A. Mills, R. Chen, J. Lee, and H. Raghav Rao, "Web 2.0 emergency applications: How useful can Twitter be for emergency response?" *Journal of Information Privacy and Security*, vol. 5, no. 3, pp. 3–26, 2009.

[14] C. A. Cassa, R. Chunara, K. Mandl, and J. S. Brownstein, "Twitter as a sentinel in emergency situations: Lessons from the Boston marathon explosions," *PLoS currents*, vol. 5, 2013.

[15] P. R. Spence, K. A. Lachlan, X. Lin, and M. del Greco, "Variability in Twitter content across the stages of a natural disaster: Implications for crisis communication," *Communication Quarterly*, vol. 63, no. 2, pp. 171–186, 2015.

[16] S. Doan, B.-K. H. Vo, and N. Collier, "An analysis of Twitter messages in the 2011 Tohoku earthquake," in *International conference on electronic healthcare*. Springer, 2011, pp. 58–66.

[17] M. Martinez-Rojas, M. del Carmen Pardo-Ferreira, and J. C. Rubio-Romero, "Twitter as a tool for the management and analysis of emergency situations: A systematic literature review," *International Journal of Information Management*, vol. 43, pp. 196–208, 2018.

[18] C. Reuter, G. Backfried, M.-A. Kaufhold, and F. Spahr, "ISCRAM turns 15: A trend analysis of social media papers 2004-2017," *Proceedings of the Information Systems for Crisis Response and Management (ISCRAM), Rochester, New York, USA*, 2018.

[19] K. Rudra, N. Ganguly, P. Goyal, and S. Ghosh, "Extracting and summarizing situational information from the Twitter social media during disasters," *ACM Transactions on the Web*, vol. 12, pp. 1–35, 07 2018.

[20] K. Rudra, P. Goyal, N. Ganguly, P. Mitra, and M. Imran, "Identifying sub-events and summarizing disaster-related information from microblogs," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 2018, pp. 265–274.

[21] M. Imran, C. Castillo, F. Diaz, and S. Vieweg, "Processing social media messages in mass emergency: Survey summary," in *Companion of The Web Conference 2018*. International World Wide Web Conferences Steering Committee, 2018, pp. 507–511.

[22] F. Alam, S. Joty, and M. Imran, "Graph based semi-supervised learning with convolution neural networks to classify crisis related tweets," in *Twelfth International AAAI Conference on Web and Social Media*, 2018.

[23] A. Hürriyetoglu, C. Gudehus, and N. Oostdijk, "Using relevancer to detect relevant tweets: The Nepal earthquake case," in *Working notes of FIRE 2016 - Forum for Information Retrieval Evaluation, Kolkata, India, December 7-10, 2016.*, 2016, pp. 76–78.

[24] J. Radianti, S. R. Hiltz, and L. Labaka, "An overview of public concerns during the recovery period after a major earthquake: Nepal Twitter analysis," in *49th Hawaii International Conference on System Sciences, HICSS 2016, Koloa, HI, USA, January 5-8, 2016*, 2016, pp. 136–145.

[25] R. Subba and T. Bui, "Online convergence behavior, social media communications and crisis response: An empirical study of the 2015 Nepal earthquake police Twitter project," in *50th Hawaii International Conference on System Sciences, HICSS 2017, Hilton Waikoloa Village, Hawaii, USA, January 4-7, 2017*, 2017, pp. 1–10.

[26] Y. Su, Z. Lan, Y. Lin, L. K. Comfort, and J. Joshi, "Tracking disaster response and relief following the 2015 Nepal earthquake," in *2nd IEEE International Conference on Collaboration and Internet Computing, CIC 2016, Pittsburgh, PA, USA, November 1-3, 2016*, 2016, pp. 495–499.

[27] M. Tsagkias, M. De Rijke, and W. Weerkamp, "Linking online news and social media," in *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM, 2011, pp. 565–574.

[28] A. Mogadala, D. Jung, and A. Rettinger, "Linking tweets with monolingual and cross-lingual news using transformed word embeddings," *arXiv preprint arXiv:1710.09137*, 2017.

[29] J. Wang, W. Tong, H. Yu, M. Li, X. Ma, H. Cai, T. Hanratty, and J. Han, "Mining multi-aspect reflection of news events in Twitter: Discovery, linking and presentation," in *Data Mining (ICDM), 2015 IEEE International Conference on*. IEEE, 2015, pp. 429–438.

[30] T. Ahmad and A. Ramsay, "Linking tweets to news: Is all news of interest?" in *International Conference on Artificial Intelligence: Methodology, Systems, and Applications*. Springer, 2016, pp. 151–161.

[31] B. Mazoyer, J. Cagé, C. Hudelot, and M.-L. Viaud, "Real-time collection of reliable and representative tweets datasets related to news events," in *Proceedings of the first International Workshop on Analysis of Broad Dynamic Topics over Social Media: BroDyn*, vol. 18, 2018.

[32] X. Lin, Y. Gu, R. Zhang, and J. Fan, "Linking news and tweets," in *Australasian Database Conference*. Springer, 2016, pp. 467–470.

[33] W. Guo, H. Li, H. Ji, and M. Diab, "Linking tweets to news: A framework to enrich short text data in social media," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, 2013, pp. 239–249.

[34] B. Shi, G. Ifrim, and N. Hurley, "Be in the know: Connecting news articles to relevant Twitter conversations," *arXiv preprint arXiv:1405.3117*, 2014.

[35] F. Abel, Q. Gao, G.-J. Houben, and K. Tao, "Semantic enrichment of Twitter posts for user profile construction on the social web," in *Extended semantic web conference*. Springer, 2011, pp. 375–389.

[36] R. McCreadie, C. Macdonald, and I. Ounis, "News vertical search: When and what to display to users," in *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2013, pp. 253–262.

[37] Z. Wei and W. Gao, "Gibberish, assistant, or master?: Using tweets linking to news for extractive single-document summarization," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2015, pp. 1003–1006.

[38] S. Kulcu and E. Dogdu, "A scalable approach for sentiment analysis of Turkish tweets and linking tweets to news," in *Semantic Computing (ICSC), 2016 IEEE Tenth International Conference on*. IEEE, 2016, pp. 471–476.

[39] H. Li and H. Ji, "Cross-genre event extraction with knowledge enrichment," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2016, pp. 1158–1162.