Identifying Post-Disaster Resource Needs and Availabilities from Microblogs

Moumita Basu*†, Kripabandhu Ghosh‡, Somenath Das*, Ratnadeep Dey*,
Somprakash Bandyopadhyay†, Saptarshi Ghosh*§
* Department of Computer Science and Technology, IIEST Shibpur, India
† Social Informatics Research Group, IIM Calcutta, India
‡ Department of Computer Science and Engineering, IIT Kanpur, India
§ Department of Computer Science and Engineering, IIT Kharagpur, India

***Abstract*—Microblogging sites like Twitter are increasingly being used for aiding post-disaster relief operations. In such situations, identifying needs and availabilities of various types of resources is critical for effective coordination of the relief operations. We focus on the problem of automatically identifying tweets that inform about needs and availabilities of resources, termed as need-tweets and availability-tweets respectively. Traditionally, pattern matching techniques are adopted to identify such tweets. In this work, we present novel retrieval methodologies, based on word embeddings, for automatically identifying need-tweets and availability-tweets. We perform experiments over tweets posted during two recent disaster events, and show that the proposed methodologies outperform the pattern-matching techniques of prior works.**

***Index Terms*—Disasters; Microblogs; Need-tweets; Availability-tweets; Word embeddings; Word2vec.**

I. Introduction

In a disaster situation, such as an earthquake or a flood, one of the primary challenges in coordinating the post-disaster relief operations is the lack of situational information. In recent years, microblogging sites like Twitter and Weibo have been shown to be very useful for gathering situational information in real-time [1], [2]. However, in such sites, critical information is usually obscured by a lot of insignificant information like opinion and sentiment (e.g., sympathy for the victims of the disaster). Since time is crucial in a post-disaster situation, automatic techniques are necessary to extract the critical information from among the content streams posted on the microblogging sites.

After consultation with members of NGOs who have worked in disaster-affected areas,[1] we identified two broad information needs which can critically help the relief efforts – *what resources are needed* and *what resources are available* in the disaster-affected area. Thus, in this study, we focus on extracting these two specific types of microblogs or tweets.

Need-tweets: Tweets which inform about the need or requirement of some specific resources such as food, water, medical aid, shelter, etc. Note that tweets which do not directly specify the need, but point to scarcity or non-availability of some resources are also included in this category. An example of Need-tweet is – *'Nepalis, r w/o water & electricity. Water is essential to be supplied to the affected people in Nepal.'* This tweet tells mainly about the need of *water* and *electricity* during the 2015 Nepal earthquake.

Availability-tweets: Tweets which inform about the availability of some specific resources. This class includes both tweets which inform about *potential availability*, such as resources being transported or despatched to the disaster-struck area, as well as tweets informing about the *actual availability* in the disaster-struck area, such as food being distributed, etc. For instance, the tweet *'@UPGovt sends 21 trucks of mineral water, biscuits and medicines to #earthquake affected #NepalQuake'* informs about the potential availability of *water*, *food* (biscuits) and *medicines* during the Nepal earthquake.

In this work, we attempt to design automatic techniques for identifying need-tweets and availability-tweets. Specifically, we develop Information Retrieval methodologies for identifying need-tweets and availability-tweets, where appropriate query-terms are selected to form queries that correspond to need and availability, and tweets relevant to the queries are retrieved. Traditionally, pattern matching based schemes have been employed by prior works for identifying specific types of tweets [3], [4]. In this work, we propose novel word embedding (Word2vec [5]) based methodologies which aim to capture the semantics of need-tweets and availability-tweets and then retrieve the tweets.

We perform a comprehensive evaluation of the proposed methodologies using tweets posted during two recent disaster events – the Nepal earthquake in April 2015, and the earthquake in Italy in August 2016. Our experiments indicate that, the proposed methodologies perform significantly better in identifying need-tweets and availability-tweets than pattern matching methodologies of prior works [3], [4].

II. Related work

In recent years, there has been a lot of work on utilizing Online Social Media (OSM) for aiding disaster relief operations [1]. However, to our knowledge, there have been only a few prior works that have specifically focused on the problem of identifying OSM posts that inform about need and availability of resources. Varga *et al.* [2] developed NLP techniques to identify such tweets. However, a large fraction of the tweets

[1]Doctors For You (http://doctorsforyou.org/) and SPADE (http://www.spadeindia.org/)

in the dataset is in Japanese, and it is unclear whether the methodology in [2] can be readily applied to tweets in English.

Some prior studies also identified patterns / lexicons which can be used to identify specific types of tweets, including tweets informing about need and availability of resources [3], [4]. To our knowledge, the most comprehensive set of such patterns has been proposed by Temnikova *et al.* [4]. We observed that a large fraction of the patterns identified in this study (referred to as EMTerms), can be used to identify need and availability of various types of resources.

Thus, the task of identifying need-tweets and availability-tweets has traditionally been approached as a pattern matching task. In the present work, we adopt a different approach – we view the tasks as *Information Retrieval* (search) tasks and propose word embedding [5] based approaches for the tasks. We demonstrate that our proposed methodologies perform better than the prior pattern matching approaches [3], [4].

III. Dataset

Microblogs related to two disaster events: For the present work, we collected tweets related to two major earthquakes that occurred in recent times – (i) the earthquake in Nepal and India in April 2015,[2] and (ii) the earthquake in central Italy in August 2016.[3] For both the disaster events, we used the Twitter Search API[4] to collect tweets that were posted during the days immediately following the event. The queries 'nepal quake' and 'italy quake' respectively, were used to collect the tweets relevant to the two events. We collected only tweets in English, as identified by Twitter itself. In total, about 100K tweets were collected for the Nepal earthquake, and about 180K tweets for the Italy earthquake.

Tweets frequently contain duplicates and near-duplicates as the same information is often retweeted / re-posted by many users [6]. Presence of duplicates can result in over-estimation of the performance of retrieval / extraction methodologies. Therefore, we eliminated duplicate and near-duplicate tweets using a simplified version of the methodologies discussed in [6]. After removing duplicates and near-duplicates, we obtained a set of *50,068 tweets* for the Nepal earthquake dataset, and *70,487 tweets* for the Italy earthquake dataset. These sets were used for all experiments reported in this study. For brevity, we will denote the two datasets as *nepal-quake* and *italy-quake* respectively.

Developing gold standards for evaluation: Evaluation of the methodologies discussed in this work required a gold standard containing the need-tweets and availability-tweets contained in the datasets. We engaged three human annotators to develop this gold standard, each of whom is proficient in English and is a regular user of Twitter, but none of whom is an author of this paper. Each annotator was given the two datasets of tweets (nepal-quake or italy-quake), and was asked to identify all need-tweets and availability-tweets in both datasets.

Each annotator was first asked to identify need-tweets and availability-tweets *independently*, i.e., without consulting the other annotators. While many tweets were identified by all three annotators in common, there were some tweets which were identified by two or only one of the annotators. Hence, we conducted a second phase, where all need-tweets and availability-tweets that were identified by at least one annotator (in the first phase) were considered. The gold standard set of need-tweets and availability-tweets were finalized through discussion with all the annotators and mutual agreement.

Finally, through the human annotation process described above, the following number of tweets were identified – *376* need-tweets and *1028* availability-tweets for nepal-quake dataset, and *177* need-tweets and *233* availability-tweets for the italy-quake dataset.

[2]https://en.wikipedia.org/wiki/April_2015_Nepal_earthquake
[3]https://en.wikipedia.org/wiki/August_2016_Central_Italy_earthquake
[4]https://dev.twitter.com/rest/public/search

IV. Identifying need-tweets and availability-tweets

In this section, we discuss methodologies for identifying need-tweets and availability-tweets. We first describe two baseline approaches and then our proposed methodology.

A. Baseline methodologies

We consider two prior studies that proposed patterns for identifying specific types of tweets posted during disaster events, including need-tweets and availability-tweets:

(1) Purohit *et al.* [3] proposed a set of 18 regular expressions to identify tweets that ask for donation of resources, and tweets that inform about availability of resources to be donated. We obtained, on request, from the authors of [3], the 18 regular expressions and use these on our dataset to identify need-tweets and availability-tweets.

(2) Temnikova *et al.* [4] proposed a large set of patterns (referred to as EMTerms) to identify specific types of tweets during emergencies. We employed three annotators (the same as those who developed our gold standard, as described in the previous section) to select those patterns which are relevant to need and availability of resources. The patterns in EMTerms are grouped into several categories, out of which the annotators identified six categories as relevant to need and availability of resources. In total, these six categories have 953 patterns. Table I shows the six categories, along with some example patterns in each category.

B. Proposed methodology

We employ Information Retrieval (IR) techniques, where the need and availability of resources are considered as information needs, and the goal is to retrieve documents (tweets) relevant to each information need. A *query* is formed consisting of *terms* relevant to an information need, and tweets containing the query-terms are retrieved. The retrieved tweets are also *ranked* based on some measure of their relevance to the query (explained later). We consider two stages in the retrieval process - first, an *initial query* is used to retrieve tweets, and subsequently, the query is expanded by adding some more terms to the initial query, and another round of retrieval is performed with the expanded query.

TABLE I
EXAMPLES OF PATTERNS FROM EMTERMS [4] THAT ARE RELATED TO NEED / AVAILABILITY OF RESOURCES (AS IDENTIFIED BY ANNOTATORS)

Category Code and Name	# Patterns	Examples of patterns
T06: Need of / offered supplies, such as food, water, clothing, medical supplies or blood	297	{Number} bags, aid, aids, bottled water, donate any supplies
T07: Volunteer or professional services needed or offered	232	volunteer heads, relief aid, help victims
C02: Needs food, or able to provide food	40	{Number} bags of rice, distributes food, donations like canned goods
C04: Logistics and transportation	232	{Number} trucks, helicopter, rescue boats
C05: Need of shelters, including location and conditions of shelters and camps	92	{Number} homeless, camps, hotel, shelter, shelter kit
C06: Availability and access to water, sanitation, and hygiene	59	need clean water, no drinking water, restoring water

Pre-processing the tweets: All tweets are pre-processed by case-folding to lower case, removal of a standard set of English stopwords, URLs and user-mentions, and stemming using the standard Porter stemmer.

Retrieval with initial query: We start with initial queries consisting of a few terms selected based on our intuition and observation of need-tweets and availability-tweets in general. For retrieval of need-tweets, we use an initial query consisting of two terms – 'need' and 'requir' (which is the stemmed form of 'require' or 'required'). For retrieval of availability-tweets, we use the initial query consisting of three (stemmed) terms – 'avail', 'distribut' and 'send'. Next, we use the following two models for retrieval with the queries:

(1) Language modeling: We employ the Indri IR system [7]. The pre-processed tweets were indexed using Indri, Then, ranked retrieval of tweets for the initial queries was done using the default retrieval model of Indri.

(2) Word embedding based model: We propose a Word2vec [5] based retrieval model that is suitable for short documents like tweets. We first train Word2vec on the tweets (of a certain dataset). For training Word2vec, the continuous bag of words model, along with Hierarchical softmax, was used, with the following parameter values – Vector size: 2000, Context size: 5, Learning rate: 0.05.[5] The Word2vec model gives a term-vector for each distinct term in the dataset which encompasses the context in which the term has been used in the dataset [5].

For a given query, we construct a *query-vector* by performing vector addition of the term-vectors of all terms in the query, and then dividing the vector sum by the number of words in the query. Similarly, for each tweet (pre-processed), we construct a *tweet-vector* by adding the term-vectors of all terms contained in the tweet and then dividing the vector sum by the number of terms in the tweet. For the query, we calculate the cosine similarity between the corresponding query-vector and each tweet-vector. We than rank the tweets in decreasing order of the cosine similarity.

Query expansion: The motivation of the query expansion phase is to add (to the query) some dataset-specific or event-specific terms, so that more relevant tweets can be retrieved. We perform query expansion using the following two methods:

(1) Rocchio expansion: We apply the well-known Rocchio expansion scheme [8] for determining the candidate expansion terms from the top-ranked tweets retrieved using the initial query. Here, after documents are retrieved using a particular (initial) query, the top-ranked k (a small number) documents are assumed to be relevant, and certain terms are selected from the top retrieved documents to expand the query. Specifically, for each distinct term in the $k = 10$ top-ranked tweets retrieved by the original query, we compute the $tf \times idf$ Rocchio scores, where tf is the frequency of the term among the 10 top-ranked tweets, and idf is the inverse document frequency of the term over the entire dataset. The top $p = 5$ terms in the decreasing order of Rocchio scores are selected for expanding the query.

(2) Query expansion using Word2vec: To expand the initial query using word2vec, we compute the cosine similarity of the query-vector with the term-vector of every distinct term in the dataset, and select those $p = 5$ terms for which the term-vector has the highest cosine similarity with the query-vector.

Note that, language model based retrieval and Rochhio expansion are traditional IR approaches, while the Word2vec based methodologies are novel methodologies proposed by us.

V. EVALUATION OF METHODOLOGIES

We now evaluate the methodologies described in Section IV. For a particular methodology, the evaluation is performed by comparing the set of tweets identified by the methodology, with the gold standard set of need-tweets and availability-tweets identified by human annotators (as described in Section III).

Note that the two baseline methodologies (described in Section IV-A) identify unordered sets of tweets, while the retrieval methodologies (described in Section IV-B) output ranked lists of tweets. We form a set of tweets by choosing the top $k = 1000$ tweets from the ranked lists, and calculate the following evaluation measures on this set.

Evaluation measures: It is important both to identify need-tweets and availability-tweets precisely, as well as to identify as many of the need-tweets and availability-tweets as possible. Hence, we use the following evaluation measures: (i) *Precision* – the fraction of tweets retrieved that are actually need-tweets / availability-tweets (according to gold standard). (ii) *Recall* – the fraction of all need / available tweets (out of all the tweets in the gold standard) that could be retrieved by a certain methodology. (iii) *F-score* – the harmonic mean of *Precision* and *Recall*. The advantage of F-score is that this measure captures the essence of both precision and recall; hence, we finally use *F-score* for comparing between methodologies.

Retrieval results: Table II shows the performance of various methodologies on the nepal-quake dataset, while Table III

[5]The Gensim implementation for Word2vec was used for all experiments – https://radimrehurek.com/gensim/models/Word2vec.html.

TABLE II
COMPARING METHODOLOGIES FOR THE **NEPAL-QUAKE DATASET**

Ranking Model	Expansion	Precision	Recall	F-score
Need-tweets				
(Baseline) Patterns from [3]		0.0075	0.0638	0.0134
(Baseline) Patterns from EMTerms [4]		0.0238	**0.8112**	0.0462
Indri	None	0.1010	0.2686	0.1468
Indri	Rocchio	0.1170	0.3112	0.1701
Word2vec	None	0.1700	0.4521	0.2471
Word2vec	Rocchio	0.1700	0.4521	0.2471
Word2vec	Word2vec	**0.1880**	0.5000	**0.2732**
Availability-tweets				
(Baseline) Patterns from [3]		0.0037	0.0115	0.0056
(Baseline) Patterns from EMTerms [4]		0.0516	**0.6337**	0.0954
Indri	None	0.3410	0.3317	0.3363
Indri	Rocchio	0.3410	0.3317	0.3363
Word2vec	None	0.4480	0.4358	0.4418
Word2vec	Rocchio	**0.4930**	0.4796	**0.4862**
Word2vec	Word2vec	0.4820	0.4689	0.4753

TABLE III
COMPARING METHODOLOGIES FOR THE **ITALY-QUAKE DATASET**

Ranking Model	Expansion	Precision	Recall	F-score
Need-tweets				
(Baseline) Patterns from [3]		0.0033	0.0904	0.0064
(Baseline) Patterns from EMTerms [4]		0.0134	**0.4576**	0.0260
Indri	None	0.0280	0.1582	0.0476
Indri	Rocchio	0.0290	0.1638	0.0493
Word2vec	None	0.0320	0.1808	0.0544
Word2vec	Rocchio	0.0280	0.1582	0.0476
Word2vec	Word2vec	**0.0650**	0.3672	**0.1104**
Availability-tweets				
(Baseline) Patterns from [3]		0.0019	0.0386	0.0036
(Baseline) Patterns from EMTerms [4]		0.0222	**0.5751**	0.0427
Indri	None	0.0210	0.0901	0.0341
Indri	Rocchio	0.0240	0.1030	0.0389
Word2vec	None	0.0400	0.1717	0.0649
Word2vec	Rocchio	0.0680	0.2918	0.1103
Word2vec	Word2vec	**0.0780**	0.3348	**0.1265**

shows the results on italy-quake dataset. We have similar observations for both datasets.

The EMTerms [4] baseline outperforms all the other methodologies significantly in terms of *Recall*. But, it achieves a very low *Precision* which accounts for its low *F-score*. These values are because the EMTerms patterns match a very large number of tweets, which include many of the actual need / available tweets (high Recall) but also match many non-relevant tweets (low precision).

In contrast, the retrieval methodologies achieve both reasonable precision as well as reasonable recall, leading to significantly better F-score values than the pattern matching methods. Especially, the proposed Word2vec based retrieval achieves the highest precision and the highest F-score values. Not only does the Word2vec based retrieval outperform the baseline pattern matching methodologies, but also the language model (Indri) based methodology.

Comparing the performance of retrieval with initial queries (the cases with no expansion) and that with expanded queries, we observe that query expansion (both Rocchio and Word2vec expansion) help to improve both precision as well as recall. For most cases, the Word2vec based query expansion, coupled with Word2vec based retrieval, achieves the best retrieval performance.

The overall superior performance by Word2vec based methods establishes the efficacy of contextual matching (where the context is captured by the term-vectors) as compared to word / pattern matching techniques, for short, noisy microblogs. We observed that, for both datasets, the Word2vec based methodologies could retrieve several need / availability tweets which could not be identified by any of the other methodologies. Table IV shows examples of some tweets which could be retrieved only by the Word2vec based methodologies. We see that these tweets talk about very specific resources (like dry foods, cooking oil, ambulances, blood etc.) and do not contain terms like 'need' or 'available' or 'distribute' (or their variants) which are intuitively associated with needs and availabilities. This is probably why pattern / word matching techniques fail to identify these tweets; however, contextual matching is able to retrieve these tweets as well.

TABLE IV
EXAMPLES OF TWEETS THAT COULD BE RETRIEVED ONLY BY THE PROPOSED WORD2VEC BASED METHODOLOGIES

Need-tweets	Availability-tweets
Nepal-quake dataset	
@ANI_newsI Requested to all News Channels Please Provide some Dry Food's, Biscuits or Mineral Waters for Nepal	#nepalearthquake Nepa Gyration(NGO), provided the earthquake victims. Rice-30 kg Cereals-40kg Cooking oil
Italy-quake dataset	
Italy earthquake Blood Donation Emergency: Below is the emergency phone number [number]	#ItalyEarthquake @WHO Field guide for crush injuries in earthquakes for medics [url]

VI. CONCLUSION

We proposed novel word embedding based techniques for identifying two types of tweets that are important for post-disaster relief operations, viz., need-tweets and availability-tweets. The proposed techniques outperform prior pattern matching based techniques.

However, the low F-scores indicate that there is still plenty of room for improving the techniques, and we look forward to exploring more effective techniques in future. We also plan to develop techniques for automatically matching resource needs with corresponding availabilities.

Acknowledgement: This research was partially supported by a grant from the Information Technology Research Academy (ITRA), DeITY, Government of India (Ref. No.: ITRA/15 (58)/Mobile/DISARM/05).

REFERENCES

[1] M. Imran, C. Castillo, F. Diaz, and S. Vieweg, "Processing Social Media Messages in Mass Emergency: A Survey," *ACM Computing Surveys*, vol. 47, no. 4, pp. 67:1–67:38, Jun. 2015.

[2] I. Varga, M. Sano, K. Torisawa, C. Hashimoto, K. Ohtake, T. Kawai, J.-H. Oh, and S. D. Saeger, "Aid is Out There: Looking for Help from Tweets during a Large Scale Disaster," in *Proc. ACL*, 2013.

[3] H. Purohit, C. Castillo, F. Diaz, A. Sheth, and P. Meier, "Emergency-relief coordination on social media: Automatically matching resource requests and offers," *First Monday*, vol. 19, no. 1, Jan 2014.

[4] I. Temnikova, C. Castillo, and S. Vieweg, "EMTerms 1.0: A Terminological Resource for Crisis Tweets," in *Proc. ISCRAM*, 2015.

[5] T. Mikolov, W. Yih, and G. Zweig, "Linguistic Regularities in Continuous Space Word Representations," in *NAACL HLT 2013*, 2013.

[6] K. Tao, F. Abel, C. Hauff, G.-J. Houben, and U. Gadiraju, "Groundhog Day: Near-duplicate Detection on Twitter," in *Proc. WWW*, 2013.

[7] T. Strohman, D. Metzler, H. Turtle, and W. B. Croft, "Indri: A language model-based search engine for complex queries," in *Proc. ICIA*. Available at: http://www.lemurproject.org/indri/, 2004.

[8] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.