Mapping Needs and Availabilities of Resource from Microblogs for Post-Disaster Relief

Wilson Nagar and K. Kalaiselvi

May 22, 2022

Mapping Needs and Availabilities of Resource from Microblogs for Post-Disaster Relief

Wilson Nagar
Computer Science Undergraduate
SRM Institute of Science and Technology, Kattankulathur
Chennai, India

Dr.K.Kalaiselvi
Assistant Professor
SRM Institute of Science and Technology, Kattankulathur
Chennai, India

ABSTRACT

During a crisis, knowing the requirements and availability of various types of resources is critical for organizing relief efforts. Microblogging sites are regularly utilized to help post-disaster relief efforts, and previous attempts have been made to find tweets that provide information on resource shortages and availability.

I. INTRODUCTION

Two forms of information are very valuable for coordinating relief operations in a catastrophe situation (e.g., earthquake, flood): what resources are needed and what resources are available in the affected area. region. Twitter and other microblogging sites are incredibly handy. During disasters, it is used to obtain situational information. As a result is a natural desire to learn more about post-disaster requirements and the data available from microblogs. There are two in particular The following types of tweets (microblogs) are beneficial: Need-to-know tweets discusses the need for or demand for specific resources, such as food water, medical assistance, shelter, and availability-tweets, which provide information about the region's resource availability, or Potential availability, i.e. resources that will be transported to the region.

II. LITERATURE REVIEW

The research paper of M. Basu (2019) focuses on the experimenting with different traditional and neural-based approaches to identify need and availability tweets including supervised and unsupervised classification approaches. These methods also involves pattern-matching techniques, CNN based classifiers and supervisor classification techniques. Overall it was concluded that unsupervised classification methodologies are more promising than supervised classification methods if the training data are rich.

Another paper based on classification is focussed on the information retrieval methodologies for identifying needs and availability tweets. M. Basu (2017) proposed novel word embedding based techniques for identifying two types of tweets that are important for post disaster i.e. needs and availability tweets which outperform prior pattern matching based techniques.

III. MODULES

A. Cleaning the dataset-

This method includes
1. Removing the user-handles, hash symbols and hyperlinks (#tread, @username, https://).
2. Removing the Unicode elements from the tweets (different Unicode symbols).

	Tweets	Label
0	RT @Economynext: #SriLanka to fly emergency medical help, food to #earthquake -struck #Nepal #lka #Economynext http://t.co/t6F2RXD4tj http:...	Avail
1	India committed in aid to Nepal: President Pranab Mukherjee: President Pranab Mukherjee... http://t.co/NeP9Pb9I0M SPS®	Others
2	RT @Thekkapoor: @narendramodi List of Doctor and hospital . #earthquake #Nepal http://t.co/pRIPlyrobc	Avail
3	Quake wake-up call for govt, need better building tech: Expert #Earthquake #Nepal #India http://t.co/85dhBJ2aES http://t.co/Q9m9eLs1ye	Others
4	#NepalEarthquake India sends 39 #NDRF team, 2 dogs and 3 tonnes equipment to Nepal Army for rescue operations: Indian Embassy in #Nepal	Avail
5	"@PMOIndia_RC: Indian Government is doing every possible help to the earthquake victims and they need money so plz contribute"	Others
6	RT @ANI_news: We have to play a leading role to provide aid to Nepal: Shashidhar Reddy (Former Vice Chairman, NDMA) http://t.co/xnzi2gaRrQ	Others
7	RT @UNFPAasia: As emergency services step up in post-quake Nepal, don't forget the urgent specialized services pregnant women need; @UNFPAN...	Others
8	1L water bottle by railways\nLocal rates from BSNL to Nepal\n250 Indians to be rescued tonight\n3.5 tones of relief material\n#HatsOffGOI	Others
9	RT @BJPLucknowBJP: Agra Medical College Building Damaged In #Earthquake http://t.co/IJvIIombVd	Others

Fig.1 Raw Dataset from the microblogs

	Tweets	Label
0	[srilanka, fly, emergency, medical, help, food, earthquake, struck, nepal, lka, economynext]	Avail
1	[india, committed, aid, nepal, president, pranab, mukherjee, president, pranab, mukherjee]	Others
2	[list, doctor, hospital, earthquake, nepal]	Avail
3	[quake, wake-up, call, govt, need, better, building, tech, expert, earthquake, nepal, india]	Others
4	[nepalearthquake, india, sends, 39, ndrf, team, 2, dogs, 3, tonnes, equipment, nepal, army, rescue, operations, indian, embassy, nepal]	Avail
5	[indian, government, every, possible, help, earthquake, victims, need, money, plz, contribute]	Others
6	[play, leading, role, provide, aid, nepal, shashidhar, reddy, former, vice, chairman, ndma]	Others
7	[emergency, services, step, post-quake, nepal, forget, urgent, specialized, services, pregnant, women, need]	Others
8	[1l, water, bottle, railways, local, rates, bsnl, nepal, 250, indians, rescued, tonight, 3.5, tones, relief, material, hatsoffgoi]	Others
9	[agra, medical, college, building, damaged, earthquake]	Others

Fig.3 After Module 2 implemented on Dataset

3. Removing stop words and punctuations.

Stop words are the English words which does not add much meaning to a sentence. They can safely be ignored without sacrificing the meaning of the sentence. For example, the words like the, he, have etc.

B. *Converting the dataset-*
   This method includes

1. Tokenization: Process of turning a meaningful piece of data, such as an account number, into a random string of characters called a token that has no meaningful value if breached. Tokens serve as reference to the original data, but cannot be used to guess those values.

2. Stemming: Process of reducing a word to its word stem that affixes to suffixes and prefixes or to the roots of words known as a lemma.

Stemming is important in natural language understanding (NLU) and natural language processing (NLP).

3. Lemmatization : Doing things properly with the use of a vocabulary and morphological analysis of words, normally aiming to remove inflectional endings only and to return the base or dictionary form of a word, which is known as the lemma

C. *Vectorization-*
   In this method, we use TF-IDF Vectorization, which helps us to transform text into usable vector. This includes-

1. Term Frequency: It is a measure of the frequency of a word (w) in a document (d). TF is defined as the ratio of a word's occurrence in a document to the total number of words in a document.

	Tweets	Label
0	: SriLanka to fly emergency medical help, food to earthquake -struck Nepal lka Economynext	Avail
1	India committed in aid to Nepal: President Pranab Mukherjee: President Pranab Mukherjee	Others
2	: List of Doctor and hospital . earthquake Nepal	Avail
3	Quake wake-up call for govt, need better building tech: Expert Earthquake Nepal India	Others
4	NepalEarthquake India sends 39 NDRF team, 2 dogs and 3 tonnes equipment to Nepal Army for rescue operations: Indian Embassy in Nepal	Avail
5	": Indian Government is doing every possible help to the earthquake victims and they need money so plz contribute"	Others
6	: We have to play a leading role to provide aid to Nepal: Shashidhar Reddy (Former Vice Chairman, NDMA)	Others
7	: As emergency services step up in post-quake Nepal, don't forget the urgent specialized services pregnant women need;	Others
8	1L water bottle by railways\nLocal rates from BSNL to Nepal\n250 Indians to be rescued tonight\n3.5 tones of relief material\nHatsOffGOI	Others
9	: Agra Medical College Building Damaged In Earthquake	Others

Fig.2 After Module 1 implemented on Dataset

$$TF(w,d) = \frac{occurences\ of\ w\ in\ document\ d}{total\ number\ of\ words\ in\ document\ d}$$

2. Inverse Document Frequency: It is a measure of the importance of a word. Term frequency (TF) does not consider the importance of words. Some words such as' of', 'and', etc. can be most frequently present but are of little significance. IDF provides weightage to each word based on its frequency in the corpus D.

$$IDF(w,D) = \ln\left(\frac{Total\ number\ of\ documents\ (N)\ in\ corpus\ D}{number\ of\ documents\ containing\ w}\right)$$

3. Term Frequency — Inverse Document Frequency (TFIDF): It is the product of TF and IDF. TFIDF gives more weightage to the word

that is rare in the corpus (all the documents). TFIDF provides more importance to the word that is more frequent in the document.

$$TFIDF\ (w,d,D) = TF(w,d) * IDF(w,D)$$

```
[[   94    71    65]
 [   27   101    49]
 [11872 20987 37221]]
              precision    recall  f1-score   support

           0       0.01      0.41      0.02       230
           1       0.00      0.57      0.01       177
           2       1.00      0.53      0.69     70080

    accuracy                           0.53     70487
   macro avg       0.34      0.50      0.24     70487
weighted avg       0.99      0.53      0.69     70487

0.5308212861946174
```

Fig.4 Results after testing the trained model

```
              precision    recall  f1-score   support

           0       0.94      0.98      0.96      1322
           1       0.93      0.93      0.93       495
           2       0.97      0.86      0.91       495

    accuracy                           0.94      2312
   macro avg       0.95      0.92      0.93      2312
weighted avg       0.94      0.94      0.94      2312

0.9420415224913494
```

Fig. 5 Results of trained model after undersampling

D. Training and Testing Dataset-

After performing TF-IDF, we continue with label encoding for the labels : Available, Needed and Others. In the current experiment, we are using Nepal Earthquake Dataset as the training dataset, and Italy Earthquake Dataset as the testing dataset to test the trained model. We train the model using GaussianNB classifier

As we see in Fig. 4, the resultant accuracy for the tested model is not quite upto the mark because the data is biased towards Others label. To remove this bias, we use Undersampling because dataset is imbalanced.

Undersampling is a technique to balance uneven datasets by keeping all of the data in the minority class and decreasing the size of the majority class. It is one of several techniques data scientists can use to extract more accurate information from originally imbalanced datasets. After Undersampling, we get an accuracy which is far better than the results before Undersampling the dataset.

REFERENCES

[1] M. Basu, A. Shandilya, P. Khosla, K. Ghosh and S. Ghosh, "Extracting Resource Needs and Availabilities From Microblogs for Aiding Post-Disaster Relief Operations," in *IEEE Transactions on Computational Social Systems*, vol. 6, no. 3, pp. 604-618, June 2019

[2] M. Basu, K. Ghosh, S. Das, R. Dey, S. Bandyopadhyay, and S. Ghosh, "Identifying post-disaster resource needs and availabilities from microblogs," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Jul. 2017, pp. 427–430.

[3] 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.vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[4] 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

[5] A. Severyn and A. Moschitti, "Twitter sentiment analysis with deep convolutional neural networks," in Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Aug. 2015, pp. 959