IIT BHU at FIRE 2018 IRMiDis Track - Obtaining Factual Tweets During Natural Disasters

Harshit Mehrotra[1] and Sukomal Pal[1]

Department of Computer Science and Engineering, Indian Institute of Technology
(BHU) Varanasi - 221005
harshit.mehrotra.cse15@iitbhu.ac.in

Abstract. This paper presents details of the work done by the team of IIT (BHU) Varanasi for the IRMiDis track in FIRE 2018. The task involved classifying tweets posted during a disaster into those which are fact-checkable or factual and which are not, and also match these tweets to relevant news articles. Methodologies had to be developed in context of the 2015 Nepal Earthquake.

Keywords: Information retrieval, microblogs, disaster, word embeddings

1 Introduction - Tasks and Data

With the increasing use of social media, the domains of its impact are also changing rapidly. In the recent past, people and media houses have resorted to social media platforms like Twitter and Facebook to post sentiments, information, need, resource availability, news updates etc. These can be a very useful source of relief relevant information. However, a lot of the information in the stream may be useless, over-stated or even contain rumors. The IRMiDis track in FIRE 2018 [1] posed the following tasks in this context:

- **Identifying factual or fact-checkable tweets:** Developing methodologies to segregate fact-checkable tweets from the huge stream of twitter microblogs to help in relief and rehab operations. Around 80 sample fact-checkable tweets are provided to develop the methodology which is later evaluated on around 50,000 test tweets.
- **Identification of supporting news articles for fact-checkable tweets:** A fact-checkable tweet is said to be supported/verified by a news article if the same fact is reported by both the media and the tweet. Each fact-checkable tweet has to be matched with its relevant news article(s) in a collection of nearly 6,000 articles. Also, the line in the article indicating the relevance has to be identified.

We submitted one run in which the methodology for the first sub-task was fully automatic and that for the second one was semi-automatic.

2 Methodology

2.1 Sub-Task 1

The methodology for the first sub-task i.e. identification of fact-checkable tweets is fully automatic in both, query generation and searching. The key steps are indicated as follows:

1. **Pre-processing** of all tweets by lower-casing, removal of stopwords, hashtags and addressing and finally stemming using porter stemmer. The term tweet hereafter refers to the pre-processed tweet.
2. Creating a **TF-IDF based ranked list** of terms in the reference set of 84 tweets. Only those terms are considered that occur in the reference set more than once. We call this set of terms R with the TF-IDF score function being T.
3. A **word2vec word embedding model** is trained on the entire set of 50,000 tweets.
4. Each test tweet is now attributed to its **corresponding feature vector** that is formed by an arithmetic mean of the sum of the individual terms embeddings.
5. To form the reference feature (V) vector against which they will be matched, we use the following weighted mean:

$$V = \frac{\sum_{i=1}^{|R|} T(R_i)E(R_i)}{\sum_{i=1}^{|R|} T(R_i)} \tag{1}$$

   E is the embeddings function.
6. Each tweet is then evaluated for its **cosine similarity** (= 1−cosine distance) with V. The similarity is normalized by dividing with the maximum similarity value obtained.
7. Now amongst the highest probability tweets, we have to separate the negative (non-factual ones). For this we form two word sets:
   (a) The first word set P is formed out of the terms in the reference dataset of 84 tweets which occur in the dataset more than once.
   (b) The second word set N is prepared as follows. Tweets having similarity less than 0.80 are taken and their terms are arranged in decreasing order of their frequency in this subset. The top 500 words in this arrangement comprise N.
8. The value 0.80 is decided by seeing the minimum similarity value of a tweet in the reference data set.
9. Since, we considered tweets with similarity less than 0.80 for negative tweets term selection, we now test the tweets with similarity greater than or equal to 0.80 against P and N. If no term of N and more than one terms of P are present in the tweet, it is classified as positive (factual).
10. The probabilities are normalized to the range (0,1] to give the factuality scores.

2.2 Sub-Task 2

The methodology for the second sub-task is manual in query generation and automatic in searching, scoring, using the Java based text search library Lucene. Details of the constituent steps are as follows:

1. The news articles are **pre-processed** in the same way as tweets are in sub-task 1.
2. The **headline and first 3 sentences** of each news articles are combined. This creates one testing document for each news article.
3. Now each pre-processed tweet is used as a query to match with the testing documents of the news articles. This done using **Lucene** and the score of the best matching document is seen for each tweet.
4. If this score is more than 0.30, the corresponding news article is said to be matching the tweet, otherwise no relevant news article is said to be found for the tweet.
5. To find the matching sentence, the tweet as a query is matched with each sentence of the relevant news article. The sentence with the highest score is returned as the answer.

3 Results

The results on the two sub-tasks, based on different metrics are indicated in Table 1 and 2.

Table 1. Results on Sub-Task 1

Rank	Run Type	Precision@100	Recall@100	MAP@100	MAP Overall	NDCG@100	NDCG Overall
5	Automatic	0.9300	0.1938	0.0709	0.1568	0.8645	0.4532

Table 2. Results on Sub-Task 2

Rank	Run Type	Precision@N	Recall	F-Score
1	Semi-automatic	0.9378	0.9756	0.9563

4 Possible Improvements

Depending on the kind of data, a sentiment analysis module can be augmented in the classification pipeline. However since such system should be ready to use for a disaster when it happens, the weight of such an additional module can be found as a hyperparameter by studying data from such incidents that have already occurred.

References

1. Basu, M., Ghosh, S., Ghosh, K.: Overview of the FIRE 2018 track: Information Retrieval from Microblogs during Disasters (IRMiDis). In: Proceedings of FIRE 2018 - Forum for Information Retrieval Evaluation (December 2018)