# Information Extraction from Microblogs Posted during Disasters

**Saptarshi Ghosh**[1]
**Kripabandhu Ghosh**[2]

[1]**Department of CST,**
**Indian Institute of Engineering Science and Technology Shibpur, India**

[2]**Department of CSE,**
**Indian Institute of Technology Kanpur, India**

# Outline

# INTRODUCTION AND MOTIVATION

# Role on Microblogs during Disasters

- Lot of useful *situational information* posted on microblogging sites like Twitter during disaster events
- Challenges in extracting the important information
  - Important information obscured amongst lot of sentiment, opinion, ...
  - Microblogs are very short and written informally
  - Large variation in vocabulary of crowdsourced content

# Motivation for the track

- Develop a standard data collection for evaluating IR methodologies for microblog retrieval during disasters
- Inspired by TREC microblog track (which does not consider disaster scenario)

# Outline

# THE TEST COLLECTION

# The Microblog dataset

- Collected tweets posted during two weeks after the devastating earthquake in Nepal in April 2015
- Used Twitter Search API with the keyword 'nepal'
- About 100K tweets in English collected
- Removed duplicates and near-duplicates based on presence of common words
- Final dataset of 50,068 tweets

# Topics for retrieval

- Consulted members of NGOs who work in disaster-affected regions – what are the typical information requirements during a disaster relief operation?
- Identified *seven* broad information requirements (*topics*)
  - FMT1: What resources were available
  - FMT2: What resources were required
  - FMT3: What *medical* resources were available
  - FMT4: What *medical* resources were required
  - FMT5: What were the requirements & availabilities at *specific locations*
  - FMT6: What were the activities of various NGOs / Government organizations
  - FMT7: What infrastructure damage and restoration were being reported

# Developing gold standard for the retrieval

- Three phases, involving human annotation and pooling
- Phase 1
  - Each annotator given the microblog collection and topics, asked to identify all tweets relevant to each topic, *independently*
  - Tweets indexed using the Indri IR system
- After Phase 1, the set of tweets identified to be relevant to the same topic by different annotators, was considerably different
- Hence, Phase 2
  - For a topic, all tweets judged relevant by *at least one* annotator considered
  - Relevance finalised through discussion among all the annotators and mutual agreement

# Developing gold standard for the retrieval (contd.)

- Phase 3 – standard pooling
  - Top 30 results of all the submitted runs pooled and judged by annotators
  - Unanimous agreement among all annotators for over 90% of the tweets
  - Majority opinion considered for the rest

# Number of tweets in final gold standard

- FMT1: What resources were available (589 tweets)
- FMT2: What resources were required (301 tweets)
- FMT3: What medical resources were available (334 tweets)
- FMT4: What medical resources were required (112 tweets)
- FMT5: What were the requirements / availability of resources at specific locations (189 tweets)
- FMT6: What were the activities of various NGOs / Government organizations (378 tweets)
- FMT7: What infrastructure damage and restoration were being reported (254 tweets)

# Examples of relevant tweets

- FMT1: What resources were available
  - India sends 39 #NDRF team, 2 dogs and 3 tonnes equipment to Nepal Army for rescue operations: Indian Embassy in #Nepal
  - If O+ve Blood is needed around Ilam, I am ready just mention. #NepalQuake
  - Dr. Madhur Basnet leading medical team going to remote villages of Gorkha dist which was epicenter of earthquake. His cell: [number]
- FMT2: What resources were required
  - Body bags, Tents, water, medicine, pain killers urgently needed in #earthquake stricken #Nepal
  - plz send medicine and food packets to nepal if possible. #NepalEarthquake
  - There is shortage of Blood as well as oxygen cylinders...Nepal is in huge crisis.
- FMT7: What infrastructure damage and restoration were being reported
  - Kathmandu-Lamjung road cut off after earthquake. Follow live updates: [url]
  - Historic Dharahara Tower in #Kathmandu, has collapsed #earthquake

# Outline

# THE TASK

# What participants were given

Participants given

- The tweet-ids, and a Python script to download the tweets using Twitter API
- The seven topics in the format conventionally used for TREC topics (number, title, description, narrative)

# Types of methodologies considered

- Automatic – both query formulation and retrieval are automated
- Semi-automatic – manual intervention involved in query formulation stage (but not in the retrieval stage)
- Manual – manual intervention involved in both query formulation and retrieval stages

# Outline

# EVALUATION

# Evaluation

- 10 teams participated in the FIRE 2016 Microblog track
- 15 runs submitted – 1 automatic, rest semi-automatic
- Primary evaluation measure – *Precision@20*; ties broken by *MAP*.

# Evaluation

| Run Id | Precision@20 | MAP | Type | Method summary |
|---|---|---|---|---|
| dcu_fmt16_1 | 0.3786 | 0.1103 | Automatic | WordNet, Query Expansion |
| iiest_saptarashmi_bandyopadhyay_1 | 0.4357 | 0.1125 | Semi-automatic | Correlation, NER, Word2Vec |
| JU_NLP_1 | 0.4357 | 0.1079 | Semi-automatic | WordNet, Query Expansion, NER, GloVe |
| dcu_fmt16_2 | 0.4286 | 0.0815 | Semi-automatic | WordNet, Query Expansion, Relevance Feedback |
| JU_NLP_2 | 0.3714 | 0.0881 | Semi-automatic | WordNet, Query Expansion, NER, GloVe, word bags split |
| JU_NLP_3 | 0.3714 | 0.0881 | Semi-automatic | WordNet, Query Expansion, NER, GloVe, word bags split |
| iitbhu_fmt16_1 | 0.3214 | 0.0827 | Semi-automatic | Lucene default model |
| **relevancer_ru_nl** | 0.3143 | 0.0406 | Semi-automatic | Relevancer system, Clustering Manual labelling, Naive Bayes classification |
| daiict_irlab_1 | 0.3143 | 0.0275 | Semi-automatic | Word2vec, Query Expansion, equal term weight |
| daiict_irlab_2 | 0.3000 | 0.0250 | Semi-automatic | Word2vec, Query Expansion, unequal term weights, WordNet |
| **trish_iiest_ss** | 0.0929 | 0.0203 | Semi-automatic | Word-overlap, POS tagging |
| **trish_iiest_ws** | 0.0786 | 0.0099 | Semi-automatic | WordNet, POS tagging |
| **nita_nitmz_1** | 0.0583 | 0.0031 | Semi-automatic | Apache Nutch 0.9, query segmentation, result merging |
| **Helpingtech_1 (on 5 topics)** | 0.7700 | 0.2208 | Semi-automatic | Entity and action verbs relationships, Temporal Importance |
| **GANJI_1, GANJI_2, GANJI_3 (Combined) (on 3 topics)** | 0.8500 | 0.2420 | Semi-automatic | Keyword extraction, Part-of-speech tagger, Word2Vec, WordNet, Terrier, Retrieval, Classification, SVM |

Table : Comparison among all the submitted runs. Runs which attempted retrieval only for a subset of the topics are listed separately at the end of the table.

# Evaluation : teamwise[1]

- **relevancer_ru_nl**: This team participated from Radboud University, the Netherlands and submitted the following *Semi-automatic* run:
  - *relevancer_ru_nl*: Run produced by a tool *Relevancer*; tweet collection was clustered to identify *coherent* clusters, manually labelled by some experts as relevant or non-relevant; Naive Bayes based classification; for each topic, the test tweets predicted as relevant by the classifier were submitted.
- **trish_iiest**: This team participated from Indian Institute of Engineering Science and Technology, Shibpur, India. It submitted two *Semi-automatic* runs described below:
  1. *trish_iiest_ss*: The similarity score between a query and a tweet is the word-overlap between them, normalized by the query length. In each topic, the nouns, identified by the Stanford Part-Of-Speech Tagger, were selected to form the query. In addition, more weight is assigned on words like *availability* or *requirement*.
  2. *trish_iiest_ws*: For this run, overlap is calculated on the synsets of each term obtained from WordNet.

[1] For the teams not presenting at FIRE 2016

# Evaluation : teamwise (contd.)

- **nita_nitmz**: This team participated from National Institute of Technology, Agartala, India and National Institute of Technology, Mizoram. It submitted one *Semi-supervised* run described as below:
  - *nita_nitmz_1*: This run was generated on Apache Nutch 0.9. Search was done using the different combination of words present in the query. The results obtained from different combinations of query were merged.
- **Helpingtech**: This team participated from Indian Institute of Technology, Patna, Bihar, India and submitted the following *Semi-automatic* run (on 5 topics only):
  - *Helpingtech_1*: For each query, relationships entities and action verbs were defined through manual inspection. The ranking score was calculated on the basis of the presence of these pre-defined relationships in the tweet for a given query. More importance was given to a tweet which indicated immediate action than a one which indicated a proposed action for future.

# Evaluation : teamwise (contd.)

- **GANJI**: This team participated from Èvora University, Portugal. It submitted three retrieval results (*GANJI_1, GANJI_2, GANJI_3*) for the first three topics only using *Semi-automatic* methodology, described below:
  - *GANJI_1, GANJI_2, GANJI_3 (combined)*: First, keyword extraction was done using Part-of-speech tagger, Word2Vec (to obtain the *nouns*) and WordNet (to obtain the *verbs*). Then, retrieval was performed on Terrier[a] using the BM25 model. Finally, SVM classifier was used to classify the retrieved tweets into *available*, *required* and *other* classes.

[a] `http://terrier.org`

# Outline

# OBSERVATIONS

# Observations

## Most used techniques

- WordNet – 8 runs, 5 teams
- NLP Tagging (NER, POS) – 8 runs, 5 teams
- Word embedding (Word2vec, GloVe) – 7 runs, 4 teams
- QE – 7 runs, 3 teams

## Possible inference

Use of external resources $\Rightarrow$ Incomplete information in tweets

## Best performances

(1 automatic, 5 semi-automatic) – WordNet (5 runs), QE (5 runs), Word embedding (4 runs), Tagging (4 runs); worst of these 5 – 15.5% better than a relatively simpler (Lucene based) method

# Outline

# FUTURE DIRECTIONS

# Future directions

- Lot of improvement in microblog retrieval still necessary
- Graded relevance, e.g., based on whether a tweet is actionable
- Tweet streams instead of tweet-set – incorporate temporal dynamics

# Acknowledgements


- Moumita Basu, Somenath Das and other annotators.
- All the participating teams.
- FIRE organizing committee for allowing us to organize this track.


Thank
You!