Annotation instruction

1. Introduction

In this task, we want to know about Hurricane Sandy by reading the tweets in 2012, the year Hurricane Sandy happened. We want to know how it happened, what was its influence on people, what people were thinking, what they exactly experienced during the disaster, what was the damage, etc.

So in this task, we will annotate tweets about Hurricane Sandy. All tweets contain keyword "Sandy", but some of them may not give us information about the hurricane. The goal of this task is to judge if the tweets are RELEVANT to Hurricane Sandy, and label the different aspects (in this task we call them TOPICs) that tweets give useful information to.

There is another important concept in this task, called ENTITY. The entity can be a city, a small area, an organization, a river, a beach, etc. Tweets about the same entity are put together, and are shown to you in time order. In the tweets group of the same entity, for any tweet in it, other tweets (especially the tweets posted before this one) can be considered as its CONTEXT. The context may help you understand a tweet.

2. Procedure

The annotation procedure includes two steps for every tweet:

First step: Judge the tweet RELEVANT or not based on whether it provides USEFUL information about Hurricane Sandy. If it is irrelevant, just choose the nethermost checkbox "Irrelevant" and pass.

Second step: If the tweet is relevant, based on the topic(s) that the tweet talks about, give it one or more TOPIC tags.

NOTICE: Pay attention to the hashtag(#XXX) and mention(@XXX) besides the plain text when you read the tweets.

2.1 How to judge if a tweet is relevant or not?

Note that the tweets containing "Hurricane Sandy", "#hurricanesandy" or other similar words/hashtags are not necessarily relevant to Hurricane Sandy. Basically, a tweet can be considered relevant if its content has ALL the following features:

➢ Related to Hurricane Sandy. The content of some tweets may have nothing to do with Hurricane Sandy, even though sometimes they contain hashtags like "#sandy".

  Examples of irrelevant tweets:
  - *Does anyone know of any good drink specials in the West Village ? #sandy #ZoneA @ Casa Tilbury http:t.coKbi5Euni*
  - *My #Mom loves weather + can't get enough of my updates . #Love her recommendation #Sandy #NYC @ West Village http:t.coOtM2GVBO*

➢ Useful. The tweet can tell you something that you care about. For example, if tweets only talk about things so personal that other people/strangers usually would not care about at all, it would be irrelevant.

  Example of irrelevant tweet:
  - *one last trip outside before sandy strikes . @ Upper East Side http:t.co6ZBapEj9*

➢ Understandable. If you cannot understand what a tweet is talking about, even after you look at its context, just consider it irrelevant. If you can figure out what it probably means, and you feel that most people, even children, could also, then it is still understandable. But if you use specialized knowledge that most people would not have, then it is not relevant.

  Example of irrelevant tweet:
  - *Take that , #sandy @ East Village http:t.coPyzrRaMG*

And here we provide some examples of relevant tweets:
  - *Stocking up at an East Village bodega because the supermarket lines are too long #Sandy2k12*
  - *ConEd called to alert that they are probably shutting power preemptively in the east village ... not stoked #sandy*

2.2 How to give a tweet topic tag(s)?

In this task, TOPIC is a subject that Twitter users often tweet about during the Hurricane Sandy event. Based on the topic(s) that the tweet talks about, give it one or more TOPIC tags.

2.2.1 Topics

We have built a typology of topics. This typology has two levels, and 9 topics that can be tagged (those have acronyms) in total.

Now we will describe every topic and give you some examples to help you understand each of them. Here is the format of our examples. The […] part before every tweet is the topic tags for this tweet. And topics are separated by stars *.

 [A*LP] " @DumboNYC : Post : Mayor Bloomberg Declares Mandatory Evacuation for Zone A (#HurricaneSandy) http:t.coFHHc2Imd " @laslugocki

t.a. = tweets about
1) People: about people's location and action before/during/after hurricane.

- **Authorities** (A): t.a. authorities, including government, police, department, and the representatives of these authorities, like mayor, president, etc. But only when these people are representing the authorities, instead of themselves.
  > Example:
  > a) [A*LP] *" @DumboNYC : Post : Mayor Bloomberg Declares Mandatory Evacuation for Zone A (#HurricaneSandy) http:t.coFHHc2Imd " @laslugocki*
  > In this tweet, Mayor Bloomberg represents the government, so topic A should be tagged.
  > b) [A] *Sanitation Dept . clearing downed trees in West Village . #sandy #HurricaneSandy http:t.copXQ7UtTc*

- **Organizations/Company** (OC): t.a. profit and non-profit organizations, e.g. red cross. E.g. their action, help, support to hurricane victims and damages.
  > Example:
  > c) [OC] *Red Cross preparing for Hurricane Sandy . Are you prepared ? Free app ! http:t.cots6RhFYN http:t.coeUEVc2BM*

- **Rescue/Relief for Sandy** (R): Rescue or relief for Hurricane Sandy, help or support for victim, relief supplies, etc. E.g. twitter users donated money to victims, or some people need support.

  Example:
  a) [R] *Help Staten Island #hurricanesandy #majorflood #help #desperate*

- **Location of People** (LP): t.a. many people's location during Hurricane Sandy; one or more people's change of location due to Sandy. E.g. people are at somewhere, not at somewhere, or evacuate from somewhere.

  Example:
  a) [LP] ***Evacuated*** *from Zone A to the Upper East Side with friends . All is well . #hurricanesandy #bettersafethansorry*
  b) [LP*S] *#NYC Times Square is **empty** ! As the city braces for hurricane #sandy . 2 hours till impact http:t.coCPDbCDXf*

- **Comment on News** (CN): People's comment on the news coverage of hurricane. This topic is NOT about the news content. Note: distinguish this topic from topic News (N).

  Example:
  a) [CN] *@AnnCurry thanks for reporting on the devastation in Staten Island caused by #HurricaneSandy*
  b) [CN] *Thank you #piersmorgan for finally some decent coverage of the devastation on Staten Island . #HurricaneSandy*

- **Death & Injury** (DI): t.a people's death or injury caused by the hurricane.

  Example:
  a) [R*DI] *My heart goes out to the mother of Staten Island who lost her two young boys due to Hurricane Sandy . My ppl of Staten Island need support*
  b) [S*DI] *Everyone needs to stop saying sandy is a pussy just because it didnt hit us hard a man in Queens died cause a tree fell on his house ! #stfu*

- **Other People related tweet** (OP): Tag this topic when you CANNOT find any other topics for this tweet, and it is **closely** related to **people**, usually people activity. So usually this topic is NOT tagged together with other topics for one tweet.

Example:
a) [OP] *Hurricane Sandy brings out the best in people . We really just peeped this , lmao . @baronkadonk @ Jersey City http:t.cowuklpHJo*

2) General

- **Situation** (S): 1) General description of the situation of the hurricane, of people, of a city, or other things. 2) The tweet is relevant, but none of other topics can be tagged.

  Example:
  a) [S] *#TheHurricane is really bad in some parts of Staten Island and Jersey . I hope everyone is safe and warm . #Sandy watching @NY1headlines .*
  b) [S] *Hurricane Sandy is doing a number on Staten Island right now .*

- **News** (N): tag it if the tweet repeats/paraphrases the content of news report. And then do not forget to find an appropriate topic(s) for the news content.
  Example:
  a) [N*S] *The newest report tells us that the storm is worse than expected . #Sandy*
  We give it topic N because this tweet is paraphrasing the news report. Then we give it topic S based on the content of the news that talks about the situation(S) of the storm.

2.2.2 Some things to notice

- Usually when you see a relevant tweet, but none of the topics can be tagged, use Situation (S).

- Multiple topic tags can be tagged to one tweet, because one tweet can talk about different topics, but usually no more than THREE topics for one tweet. Here is some statistics in gold-standard corpus, 15% of tweets are irrelevant, 76% one topic, 8% two topics, 1% three topics.

- The content of the tweet may be limited, so when what the tweet is talking about is not so clear, you can do some reasonable inference based on the context and your background knowledge, but NO excessive association and

imagination. Also notice that: the background knowledge you use MUST NOT be specialized knowledge; use the knowledge that you think most people have.

- You may just see PART of all the topics during your annotation.

- Remember that **negation** does NOT affect the topic tagging. For example, for tweets "Thank god! Nobody is injured." and "Someone is severely injured in Sandy", topic DI should be tagged to both of them.

2.2.3 Topic Typology

Here is the typology of all topics:

❖ People
  ▪ Authorities (A): e.g. government, police, department, mayor, president
  ▪ Organizations/Company (OC): e.g. red cross.
  ▪ Rescue/Relief for Sandy (R): rescue, relief, help, support. E.g. relief supplies, donation, asking for help.
  ▪ Location of People (LP): 1) location of people in Sandy 2) change of location due to Sandy. E.g. evacuation, empty.
  ▪ Comment on News (CN): NOT the news content. People's comment on the news coverage.
  ▪ Death & Injury (DI): death or injury caused by Sandy
  ▪ Other People related tweet (OP)

❖ General
  ▪ Situation (S): 1) general description 2) tweet is relevant, but no other topics can be tagged here
  ▪ News (N): tweet repeats/paraphrases the content of news report