

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 8 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 *.

[PW*C] Dear friends and family , I'm alright . Power still out on the West Village and I have zero AT & T connectivity . #Sandy

t.a. = tweets about

1) The Influence of Hurricane on Human society: about the influence of hurricane on different aspects of human society.

- **SuPplies (SP)**: t.a. **articles for daily use that people buy on their own according to their needs**. E.g. **food, clothes, wood**, etc. The places to buy them can be **stores, supermarkets, restaurant**, etc. Note that **relief supplies are NOT included in SP**, because it should be things that people buy on their own.

E.g. people stock up supplies before hurricane, and the tweets asking where to buy food after hurricane...

Example:

a) [SP] *Stocking up at an East Village bodega because the supermarket lines are too long #Sandy2k12*

- **PoWer (PW)**: t.a. power and electricity condition.

Example:

a) [PW] *#moon watcher **#blackout** #nyc #sandy @ East Village <http://t.coQwt05wRj>*

- **Water supply (W):** Water supply condition in the city
Example:
a) [W] *What 's the water situation in East Village downtown ?
Worked fine this morning but heard people say there is none
now ? #Sandy*
- **Road patency/Traffic (RT):** t.a. whether **roads/streets/bike trails/etc.**, are **clear or blocked**, **traffic flows** in the city, etc.
Example:
a) [RT] *Traffic getting to normal on the Upper East Side the
Day After Sandy @ Upper East Side <http://t.co/Klg73iec>*
b) [RT*PW] *Walking through West Village & NYPD all
over patrolling - trees & scaffolding down - no power-
lots of streets blocked to vehicles- #Sandy*
- **Cell service (C):** t.a. cell service condition in the city. E.g. **telephone signal**, **network signal**, **AT&T**, **Verizon**, etc.
Example:
a) [PW*C] *Dear friends and family , I'm alright . Power still
out on the West Village and I have zero AT & T
connectivity . #Sandy*
b) [C] *Trying to hear Mayor Bloomberg speak and my Fios
keeps freezing ...*
- **Public Facilities (PF):** DIRECT description about **whether public facilities** (like school, park, subway, etc.) are **open or closed**. Do NOT give tweets this tag just because the content of tweet is somewhat related to public facilities.
Example:
a) [PF] *Mayor Bloomberg announced he 's turning off the
subways at 7 ... Good luck tonight NYC ! #hurricanesandy
#Northeast #prayers*

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 .

- **Scene (SC):** The tweet is posted **ONLY** to show a scene of a place with a picture URL, but the user **does not give any clear textual description of the picture**. Therefore, you **CANNOT** know what the scene looks like.

SC tweet is **relevant but not clear and useful**, so we **may treat it as irrelevant** in some processing, so **choose this topic CAREFULLY**.

Structure feature of SC tweet: it **must contain URL**; and **the text is often shorter than one sentence**. A common structure: (here is) + **location name + URL**.

Example:

- a) [SC] Here 's what the East Village looks like right now . Corner of Ave C & 8th Street . #Sandy http:t.cowqCNNcX

After reading this tweet, we do not know what the East Village looks like, but we know it is relevant, so we use SC.

- b) [SC] Hurricane Sandy #hudson #jersey @ Hudson River http:t.cotEyimx6T

- c) [SC] More storm videos from the Bronx #Sandy http:t.coCL5SBGmz

- d) **[OC]** Red Cross van on Delaware Ave #Buffalo #hurricanesandy #frankenstorm http:t.coYNDbIvJA

This tweet is used to show a scene, too. But the text gives us some useful information and thus we know what the scene looks like. So **DO NOT give this tweet SC tag**.

- e) **[S]** Video report : Battery Park , NYC , Hurricane Sandy on the way http:t.coyAAk8gx9 #Sandy #sandyNY

This tweet gives us useful information (the underlined part). So we **give it S tag instead of SC**.

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, 16% of tweets are irrelevant, 78.5% one topic, 5% two topics, 0.5% 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 “I am happy that we still have power.” and ”Oh no! Power off.”, topic PoWer should be tagged to both of them.

2.2.3 Topic Typology

Here is the typology of all topics:

- ❖ The Influence of Hurricane on Human Society
 - SuPplies (SP): buy food, clothes, wood, etc., from stores, supermarkets, etc. Eg. Stock up with food
 - PoWer (PW): power/electricity condition. Eg. Power off, #blackout
 - Water supply (W)
 - Road patency/Traffic (RT): roads/streets/bike trails/etc., are **clear or blocked**; condition of traffic flows...
 - Cell service (C): telephone signal, network signal, AT&T, Verizon, etc.
 - Public Facilities (PF): school, park, subway, etc., are **open or closed**
- ❖ General
 - Situation (S): 1) general description 2) tweet is relevant, but no other topics can be tagged here
 - Scene (SC): to show a scene; must contain a URL; but text part of the tweet does NOT give much useful information about Sandy