Human Sensation Dataset :

Analyzing social media have appreciated in value for understanding human behaviors to grasp public interests or emotions, as both the medium and outcome of human experiences. In particular, human emotions, regarding both physical and linguistic, are mostly dependent on sensory perceptions according to psychology and neuroscience. Even though sensation is the most fundamental element as an antecedent of human emotions, the rack of background resources makes it hard to study the social sensation in automatic fashion comparing with the sentiment or emotion analysis. The dataset is for the research of human sensation information in terms of natural language. This is the first knowledge for analyzing human sensations categorized five sensation types, such as sight, hearing, touch, smell, and taste.

Data Source :

WASSA-2017 Shared Task on Emotion Intensity (EmoInt) Part of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA-2017), which is to be held in conjunction with EMNLP-2017.

Task :

In order to extract human sensory experience as well as discover an evidence of sensation effect on emotion, the dataset is annotated again in terms of sensation types based on the crowdsourcing platform (MTurk). The tweets in EmoInt were then shown to 5 annotators who rated them in terms of five kinds of sensations using the 4-point Likert scale "0 (definitely not)", "1(perhaps not)", "2 (perhaps yes)", and "3 (definitely yes)". To improve the evaluation quality following our purpose, the annotators needed to (a) pass a qualification test with 70% accuracy, (b) reside in an English-speaking country, and (c) had high credibility from the previous tasks they contributed to. We have decided to directly accept a result if it was annotated with the same decision by three out of five individuals. If there was no agreement from at least three annotators, then we used the majority value from the five values (however, almost two or more).

Data Structure :

The dataset is annotated from emotional tweets as five types of sensations: sight, hearing, touch, smell, and taste, considering text, emoticon, and hashtag. Based on the annotation result, we separate the dataset as two categories: Binary sensation set and Multi-sensation set.

Binary-sensation_set :

contains two classes, i.e., sensation or not, with label and tweets. We classified tweets based on the annotation result by the rule that classify tweets which have greater than 1 (perhaps yes) of the agreement. Otherwise, we classify tweets to the none-sensation class

Summary Table :

Tweet	Label
#amwriting \#horror in the dark and a loud creaking door noise is coming from the kitchen. There are no doors there to creak. WTF	Sensation
When a guy comes on the train that smells like a mixture of a damp dog, old sweat and sewage works!!!! #gross#getoffthetrain #smelly	Sensation
How is it suppose to work if you do that? Wtf dude? Thanks for pissing me off.	none
honestly all my friends do bnc i mean it and i am feeling #terrible hahaha	none

Multi-sensation_set :

includes six classes, i.e., sight, hearing, touch, smell, taste, and none, with label and tweets. The tweets are classified to the six classes as same as the rule of Binary-sensation class. Thus, if a tweet has an annotation value as greater than 1 of each sensation, then it will be determined to the sensation class. Note, some tweets can have multiple sensations, such as sight and hearing, hearing and touch, and so on; In this case, we duplicated the tweet and assigned to the corresponding classes.

Tweet	Label
so if whichever butt wipe pulled the fire alarm in Davis because I was sound asleep	hearing
@li** I had a nice Italian ice-cream whilst resting my tired paws. Honey flavoured, naturally!	taste
Someone let snakes in my house, I bet it @Yt** I kill that bugger when I get my hand on him \#rage	sight/touch
Sometimes the worst place you can be is in your own head.	none

Statistics :

Distribution table of emotion and sensation variables

Emotion Intensity	Sensation	None-sensation
Strong Intensity (Emotion\>0.5)	756	453
Normal Intensity (Emotion\<0.5)	2270	1230
None (Emotion=0)	1594	1796

Proportion of tweets annotated with a score higher than 1

Sensation	Proportion (%)
Sight	78.03
Hearing	28.17
Touch	11.59
Smell	1.04
Taste	1.42

P values from chi-square test between all emotions and sensation types

	Sight	Hearing	Touch	Smell	Anger
Anger	2.14e-09*	2.2e-16*	0.00044*	0.1825	0.0580
Fear	2.2e-16*	9.72e-13*	1.65e-1*	0.8213	0.0199*
Joy	2.2e-16*	2.2e-16*	1.54e-10*	0.0481*	0.6735
Sadness	0.08866	0.04374*	0.00357*	0.5306	0.7630