Analyzing User Sentiments on Weather-Related Tweets

Nirmal Kanagasabai

McGill University, Canada

Abstract

The purpose of this exercise is to analyze user sentiments on weather-related tweets. As the task is strenuous and time-consuming, CrowdFlower, a Crowd-sourcing platform was made use of to ease the analysis aspect of the exercise. Data Collection was done using Twitter's Streaming API, data cleansing was carried out and the dataset was uploaded to CrowdFlower. To achieve the goal, two different questions were posed per tweet to the contributors. The first question was to analyze if the tweet was related to weather and if yes, what was the user's sentiment (positive, neutral and negative). We were particularly interested in inferring what weather extremity (hot, cold, wet and dry) did the users have an aversion to. So, the second question to the contributor was to infer from the tweets what weather extremity the user was referring to. The job was designed and several considerations were made during this phase. Clear instructions with examples and validation questions were provided to the contributors to assist them in carrying out the task. The enriched data was downloaded from the platform and the inferences were made accordingly.

Keywords: Sentiment-Analysis, Weather-Related Tweets, CrowdFlower

1. Introduction

Sentiment analysis have been one of the most common inter-disciplinary research areas. In this exercise, we focused on analyzing user sentiments based on their weather-related tweets. There has already been good research works conducted in this area [1]. The idea is to experiment and understand the user sentiments expressed in the Tweets and also to figure out which weather extremity he/she had an aversion to. Several sub-tasks were carried out to achieve the goal.

2. Data Collection

A lot of research has been carried out to analyze the user sentiments on different regards (Weather, Politics, etc.) based on the users' on-line footprint. To taper such sentiments, the best and the most available public platform 'Twitter' was made use of in this exercise. Using Twitter's Streaming API, a lot of tweets related to 'weather' was collected and cleansed to suit the purpose. The weather updates (from the Twitter handles of Meteorological stations) did hinder the process. Data cleansing had to be carried out so that only the Tweets posted by users where they expressed their sentiments were retained. Only the Tweet content was scrapped and the dataset was prepared.

3. CrowdFlower

3.1. Designing Phase

A new job was created from the scratch and the raw dataset that was prepared was uploaded to CrowdFlower. In the designing phase, the column 'Tweets' were made use of as the Data. Initially, three questions were created. The first question was to check if the Tweet was related to weather or not. However, as the job of the contributor increases with the increase in number of questions and eventually, the cost also increases, a design decision was made to cut down the number of questions to only 'Two' and that too, the second question would pop up only for certain circumstances. The following multiple choice questions were created and posed to the contributors:

1. What is the author's sentiment (feeling) towards the weather?

   - Positive
   - Negative
   - Neutral / The author is just sharing the information
   - This tweet is not related to weather

2. Is the author referring to any specific weather extremity? If yes, please select the appropriate one.

   - Hot
   - Cold
   - Wet

- Dry

- The author hasn't mentioned anything explicitly

At first, only the first question is posted to the contributor. If the author's sentiment is captured as 'Negative', the second question pops up asking the contributor to identify which weather extremity was the author referring to. The conditional logic was applied so that the second question gets triggered if the response to the first question is 'Negative'.

3.2. Instructions

Clear and detailed instruction set was prepared and was added to the CrowdFlower platform. While working on the instructions, there was a need to revisit the different Tweets that was collected to understand how the authors have expressed their sentiments. Based on the inference, the instruction set to classify the tweets into different classes was prepared. For the second question, the instruction set was inadequate. Originally, while creating the multiple choice, it was provided as 'Hot / Humid', 'Cold / Freezing', 'Wet / Precipitation' and Dry. However, while previewing the Tweets, it was obvious that there were many more synonymous words that the users used to define what the weather was. Hence, a decision was made to switch the multiple choices from 'Hot / Humid' to 'Hot', 'Cold / Freezing' to 'Cold' and 'Wet / Precipitation' to 'Wet'. Also, a new section called 'Glossary' was added to the instruction set to define what the weather extremity corresponds to and what other synonymous words the authors can make use of.

3.3. Examples and Validation Questions

To make things more illustrative, 10 examples were created (attributing to all possible categories). After creating the instructions, the 'quality' phase of the task was visited. Here, for 100 units (rows in the dataset), a minimum of 8 validation questions had to created. Also, I had to balance the 'Answer Distribution' by answering at least one question to each category. Hence, I created 9 validation questions (instead of 8) ensuring that all categories are covered.

3.4. Cost Considerations

There was another consideration on not having too many questions and a bigger dataset as there were budget limitations. For each tweet, as it is quite tricky to understand the sentiment, contributors would be paid at 2 cents per judgment. Hence, increasing the number of questions will increasing the number of judgments made and eventually, the cost rises.

3.5. Problems Encountered (Post assessment from Contributor)

The job was launched internally (for cost-effective purposes) and the URL was shared to the contributor (only one in this case). However, the contributor reported an issue back to me after attempting to take up the task. Due to my design flaw and the amateur understanding of how CrowdFlower portal works, the contributor had to take up the task using 11 different accounts (only to be expelled 10 times in the validation phase and once during the first set of tasks). I decided to look up where the contributor was going wrong and what the mistake was at my end.

3.6. Understanding the Design Flaw

The full report was downloaded from CrowdFlower for all the 11 occasions. In total, there were 55 tweets (most of them were duplicates and repeated on several occasions). There were 21 unique tweets (17 of them were golden units and 4 of them werent). As there isnt any way to validate the 4 tweets without looking up on it manually, I decided to focus on the 17 tweets which were golden units.

In those 17, 3 Tweets were marked wrong (on 9 occasions) despite the contributor providing the right answers. The reason behind this validation error is my decision to change the answers from Hot/Humid to Hot, Cold/Freezing to Cold and Wet/Precipitation to Wet. I assumed that CrowdFlower is going to match the changes in the data-set. However, to my shock, the answers (Hot, Cold and Wet) were treated as additional options to the question instead of replacing the original ones.

I created pivot tables to count the number of occurrences of each tweets, the count of correct answers and the count of incorrect answers based on

4

the golden units. Because of my design error, the 3 Tweets (which were occurring on 9 different occasions) were counted as incorrect answers. Leaving that apart, out of 42 other validation questions, the contributor answered correctly on 29 occasions (constituting 68.05%) and wrongly on 13 occasions (which constitutes 31.95%) as depicted in Figure 1.
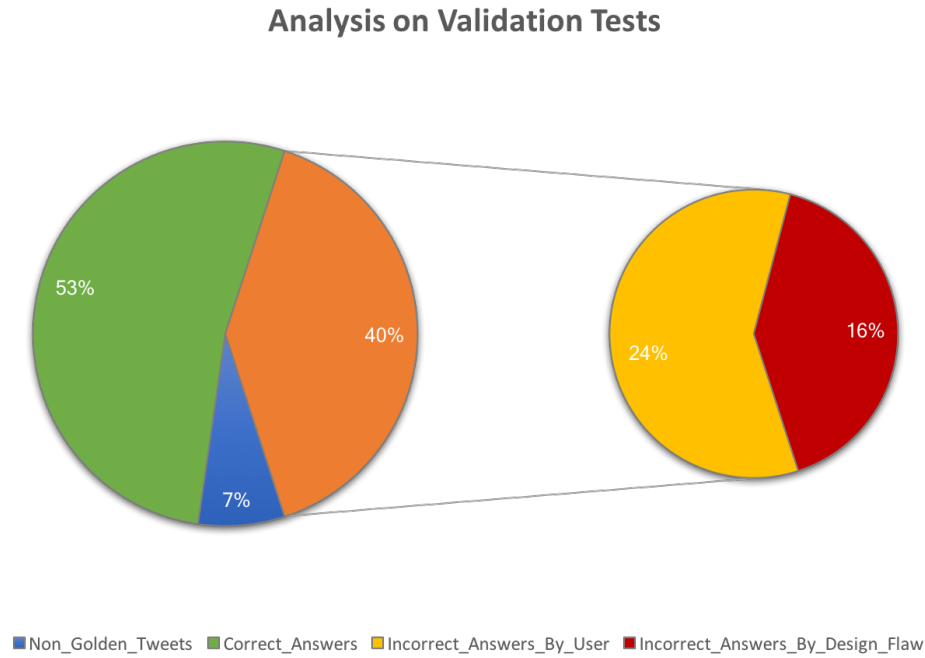
Analysis on Validation Tests

Figure 1: Validation Tests

From Figure 1, it can be observed that the incorrect answers were not only because of my design flaw but also the misunderstanding of the instructions by the contributor. One important observation was made while discussing the same with the contributor. She was merely using the keywords provided in the Glossary to predict the sentiment and was not looking at the contextual meaning of the words. Hence, few mistakes were carried out by the contributor. Clarifications were done and after making the necessary changes and relaunching the job, she was able to achieve 100% in the tasks.

4. Results and Discussion

The enriched data was downloaded from the CrowdFlower portal and inferences were made using the columns populated by the contributor. As there was an upper bound on the number of rows (units) she could answer, the rest of the units were worked by me to make the inference possible. This was done to ensure that the dataset was not super-small. This was also because of the constraints on the budget and time.

While cleansing the data, a lot of non-weather related tweets were removed and hence, it contributed to only 3% of the total number of tweets. Unfortunately, the dataset was heavily skewed with two categories (Negative and Positive sentiments) contributing to 92% leaving behind 5% for the Neutral sentiments as seen in Figure 2.

As the negative sentiments were constituting 49%, I wanted to study which weather extremity the author was displeased with. 18% of the Tweets were generic and the author just stated the keyword 'weather' not explicitly mentioning what the weather was back then. A total of 22% was taken up by 'Wet' and 'Hot' conditions (11% each). It was closely followed by '9%' where the author found 'Cold' weather to annoy them leaving behind 2% for 'Dry' weather.

5. Conclusion

To summarize, the CrowdFlower platform was of great use to lessen the efforts of the research community outsourcing the data enriching process to random contributors across the globe. Also, it servers as a great utility to split down the original task into small micro-tasks and thereby validating if the contributor is actually serious about the task he/she is carrying out. Also, it makes it easier for the research community to just focus on using Machine Learning algorithms to better understand the enriched data and not worrying about outsourcing the work to serious contributors and calculating the finances (which is carried out by CrowdFlower).

It is to be observed that the results of this exercise matched with the reference paper [1]. Considering the tight budget and time constraints, the work
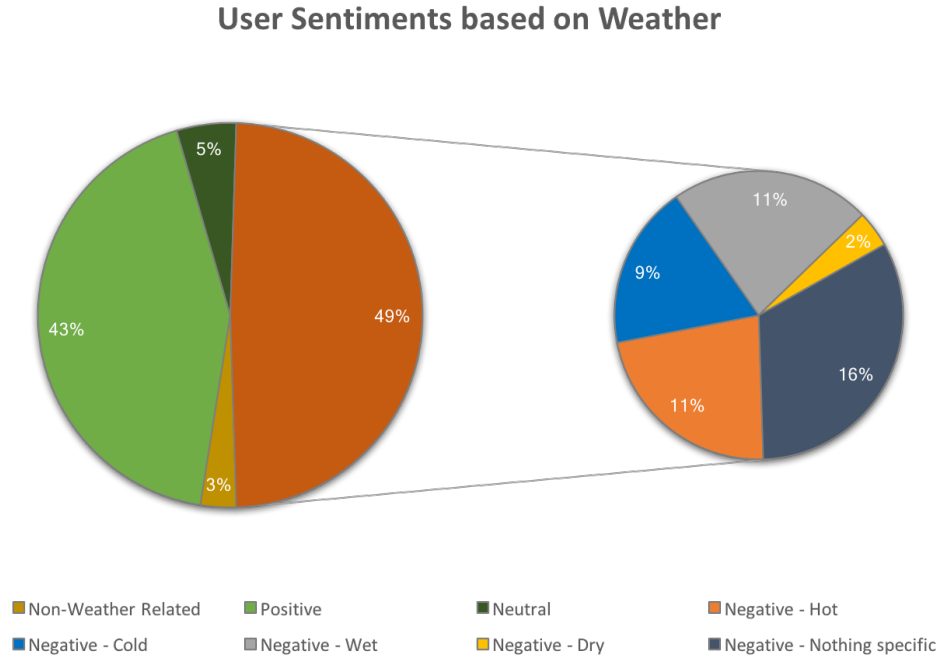
Figure 2: User Sentiments

was limited by a small dataset and only one contributor. The future works will involve larger dataset scraped across multiple platforms (Twitter, Facebook, etc.) and more contributors to enrich the dataset in the CrowdFlower platform.

6. References

[1] Patrick Baylis, Nick Obradovich, Yury Kryvasheyeu, Haohui Chen, Lorenzo Coviello, Esteban Moro, Manuel Cebrian, James H. Fowler,"Weather impacts expressed sentiment", Computation and Language (cs.CL), arXiv:1709.00071, Cornell University Library, 31 Aug 2017.