Project Proposal
CIS 400
Hong Yang Chen, Michael Jones, Robert Robinson, Hunter Malley, Julia Barucky
Product Name: Weatherments™

Project Idea

For our project we plan to use sentiment analysis on location specific tweets combined with data on the weather of that location to try to determine the general trend of how weather affects people's moods and how they translate into tweets. We will be focusing on the Syracuse area; however, the area of focus can be changed by the user.

Sentimental analysis "also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions " (Professor Edmund Yu's slides); this will be the main driver of our research. We will use nltk to determine the sentiments of tweets which can then be aligned with weather sentiments. To get the weather sentiments, we will create our own mapping from a combination of various weather data points (temperature, cloud coverage, precipitation, etc.) to one of three sentiments about the weather: good, ok, or bad. A good day is exceptionally good with little cloud coverage, temperate temperatures, and no precipitation. A bad day is exceptionally poor with heavy cloud coverage, heavy precipitation, or undesirable temperature extremes.

Features

- Analysis of the sentiment of tweets and how they correlate to the weather at the time of the tweet.
- Select different locations to focus in on
- Determine whether a tweet is a positive tweet or a negative one
- Determine if the weather is good, ok, or bad
- Show a correlation between weather and mood (if there exists)
- Determine which weather is best for peak interaction on a tweet (i.e. in what weather will someone get the most likes or retweets)
- Determine some popular topics in a given weather

Significance of Idea

This idea is significant because we hope to be able to get insight into how the weather affects people based on data pulled from twitter. This data can be broadened to look into disorders like

seasonal depression and other environmentally caused issues. Additionally, this data can be analyzed to determine what weather results in increased activity and engagement. In a sense, our output would act like a massive, and public, journal capturing moods and correlating them to more specific weather events or types of weather. Since we cannot run the program for whole seasons at a time, we will rely on timeblocks of different weather to provide us with data for making our end game correlations. It will look at not seasonal affective disorder, but more specifically the weather in a given period smaller than a season. Our idea is important because it is attempting to shed light on the abstract and difficult to pin-point problem of environmentally influenced moods.

Work Plan

Week1 (All to be finished by Sunday 4/18) - Week of Research

- Research on what algorithm and modules to use (Everyone)
- Github repository (Mike)
- Determine what database work well with python and data mining (Mike)
- Research on different weather APIs (Julia, Hong)
- Research python packages needed (Hunter, Rob)

Week2 (All to be done by Sunday 4/25) - Week of algorithm implementation

- Coding module to database connection (Mike)
- Implementing weather APIs (Julia, Hong)
- Implement streaming api to retrieve tweets (Hunter, Hong, Rob)

Week3 (All to be done by Sunday 5/2) - Week of collecting data

- Create dictionary mapping weather data to sentiments (Hunter, Mike)
- Use scraping algorithm to collect twitter data (Rob)
- Determine average sentiment of tweets in a given period (Rob, Hong)

Week4 (All to be done by Sunday 5/9) - Week of Presentation

- Comparing tweet sentiments to weather sentiments (Julia, Rob, Hunter)
- Comparing weather sentiments to total tweet numbers(Mike, Hunter, Hong)
- Presentation/ Gathering final outputs (Everyone)

Week5 (To be done by 5/16)

- Final report and submission (Everyone)