

Emojis and Weather

Darryl Hannan
Villanova University

Emma Chin
Villanova University

Matt Mador
Villanova University

December 10, 2017

Abstract

A significant amount of research has been done analyzing the impact weather has on emotional states. Most of these studies are survey based, and therefore have a small sample size. Here, we seek to overcome this problem through the use of social media. We analyze tweets from various geographic locations, extracting emojis. We use the distribution of emojis to represent the emotional state of a population. We use these representations to train a machine learning classifier that predicts the weather. Our model is approximately 88% accurate, providing support for the link between weather and mood.

1 Introduction

The impact that weather has on mood is well established. Spasova (2011) found that abrupt transitions to unfavorable weather conditions has a negative effect on emotion.[5] Keller et al. (2005) found that time spent in pleasant weather results in better mood and improved memory.[2] Some effects are more serious; seasonal affective disorder occurs during winter months, when the days are shortest, and can result in serious cases depression.[3]

The effect that weather has on emotions varies significantly among individuals.[5] Therefore, large sample sizes are needed to discover correlations between specific weather patterns and emotional states. Traditionally, the only way to

study this set of phenomena was through survey based methods, resulting in small sample sizes. We attempt to overcome this problem through the use of Twitter. By analyzing the emotions expressed in tweets for a given geographic location, we obtain an approximation of the collective emotional state of the population. This method also makes it significantly less expensive to collect data. By running the collection program on a server, we can continuously collect data indefinitely.

2 Related Work

Park et al. (2013) also utilize Twitter to gather emotional data.[4] They use a linear regression model to predict various variables such as temperature, humidity, and precipitation. Our approach differs in that Park et al. perform sentiment analysis on the tweets, whereas we use emojis to represent sentiment. By using emojis, we require significantly less preprocessing and impose less structure on the emotional representation that the model ultimately learns.

Baylis et al. (2017) use both Twitter and Facebook to collect emotional data. Using more than 3.5 billion social media posts in their model.[1] Their weather variable, referred to as cloud cover, is similar to the variable that we predict in our model. The primary difference in our model again lies in the use of emojis.

3 Data

3.1 Emotion Data

We collect tweets from 21 cities using Twitter’s Streaming API via Tweepy¹, targeting locations with high populations in order to maximize the number of data points. We preselected a list of 75 emojis that commonly represent various emotional states (Figure 1). As each tweet is received, we count the occurrences of each emoji in the tweet. We collect data in 2 hour intervals. The distribution of emojis for each interval is represented in a vector format, where each dimension of the vector contains the count of a given emoji. Data was collected over a period of approximately 2 weeks.

¹<http://www.tweepy.org/>

1	😄
2	😄
3	😄
4	😄
5	😄
6	😄
7	😄
8	😄
9	😄
10	😄
11	😄
12	😄
13	😄
14	😄
15	😄
16	😄
17	😄
18	😄
19	😄

Figure 1: Example of emojis that we extract from tweets

3.2 Weather Data

We access weather data for each location through the OpenWeatherMap API. We use PyOWM², an open-source python wrapper, to simplify API calls. Weather data is collected concurrently with Twitter data. At the end of each 2 hour interval, we collect the current atmospheric condition in each city. Examples of atmospheric conditions include rain, snow, clear, etc. In the data we collected, there are 8 classes of conditions: clouds, clear, rain, drizzle, mist, fog, snow, haze, and dust.

3.3 Preprocessing

One issue that we encountered was that some of the emoji vectors were extremely sparse. In some cities, such as Nashville, almost every vector only had a few emojis. Therefore, we only include vectors in which the sum of its elements is greater than 40. For classification, each atmospheric condition is represented using a number from the range 0-7. In the case of

²<https://github.com/csparpa/pyowm>

binary classification, we consider clear to be a positive weather condition and represent it with a 0. We consider all other conditions to be negative and represent these using a 1. Our final data set consisted of 977 data pairs.

4 Results

In this section we describe the results.

5 Conclusions

Conclusion goes here.

References

- [1] Patrick Baylis, Nick Obradovich, Yury Kryvasheyev, Haohui Chen, Lorenzo Coviello, Esteban Moro, Manuel Cebrian, and James H. Fowler. Weather impacts expressed sentiment. aug 2017.
- [2] M. C. Keller, B. L. Fredrickson, O. Ybarra, S. Cote, K. Johnson, J. Mikels, A. Conway, and T. Wager. A Warm Heart and a Clear Head: The Contingent Effects of Weather on Mood and Cognition. *Psychological Science*, 16(9):724–731, sep 2005.
- [3] Sherri Melrose. Seasonal Affective Disorder: An Overview of Assessment and Treatment Approaches. *Depression Research and Treatment*, 2015:1–6, nov 2015.
- [4] K Park, S Lee, E Kim, M Park, J Park, and M Cha. Mood and weather: Feeling the heat? *7th International AAAI Conference on Weblogs and Social Media, ICWSM 2013*, pages 709–712, 2013.
- [5] Z Spasova. The effect of weather and its changes on emotional state individual characteristics that make us vulnerable. *Adv. Sci. Res*, 6:281–290, 2011.