CSC 240 Term Project Proposal
(Happiness and Partnership)

Chester Holtz

1 Introduction and Problem

Defining happiness can seem as elusive as achieving it. Humans want to be happy, and we can say whether we are or not, but how can we approach formally defining or measuring happiness such that a machine can understand happiness? Can we use this definition and understanding to become happier?

Additionally, common psychological research also distinguishes between the moment-by-moment feeling of happiness produced by positive emotions and how we describe our lives when we think about it. Regardless of whether you one has a happy day or not, describing one's life as a happy one is more difficult - or can one describe themselves as a happy person? Psychologist Daniel Kahneman describes this difference as the experiencing self and the remembering self. Psychologists study both to better understand how daily experiences add up to a happy life

Psychology researchers have formed a useful framework for studying happiness:

- Happiness is made up of pleasure, engagement, and meaning
- It involves both daily positive emotions and a global sense that life is worthwhile
- People can accurately report their own levels of happiness

[8]

In particular, this study will examine happiness from the perspective of computer science. Is there a way we can examine someone's digital behavior and quantify their happiness based on their social media activity - primarily their interests, hobbies, and especially the level of companionship they experience.

2 Related Work

Sentiment analysis of social media data is a well-studied issue. Even effective metrics of happiness for artificial learning agents is also a prior studied problem. To the best of my knowledge, use of a centralizing service such as about.me to mine and evaluate the data of a user based off of the content of several of his social media profiles is not as well studied.

2.1 Methods

A thorough review of the common algorithms utilized by approaches discussed in the related works is outside the scope of this paper. This paper will primarily be concerned with developing a model to discover some correlation, and potentially predict, happiness and companionship, religion, race, and other factors. However, we will give a high level overview of the most popular algorithms.

There are many effective algorithms we can leverage to preform tasks such as sentiment analysis, classification, and pattern/rule mining. Common algorithms utilized by approaches earlier discussed to preform sentiment analysis are Bag of Words and Naive Bayes (utilized by Hedonometer)[3], SVM and Log-Linear Regression Models (Volkova's Work)[7], and even crowd-sourcing such as Amazon's Mechanical Turk[4]

2.2 Critical Evaluation

Daswani et al. presents a formal definition of happiness in the context of machine learning. Additionally, he describes a resulting Markov Decision Process of unhappy and happy agents based off of his formula. He presents several examples and does analysis on the differences in classification based on his model and empirical research done by trained psychologists. In particular, he partitions *happiness* into 4 desired properties: scaling, subjectivity, commensurable, and agreement.

Dodds et al. provides the most relevant work to this project in which he implements a model of happiness prediction based off of tweets as a web application he calls a *Hedonometer*. Namely, the Hedonometer also incorporates an extensive dataset of classified words and happiness weights. Unfortunately, one criticism I have for this work is that the papers seemed to have been produced as a secondary thought to the app, and no real research has been conducted on the method of happiness classification - ie Dodds et al. classifies a person as "happy" if the average sentiment of their tweets exceeds a certain threshold.

Cody et al. provides a successful example of actual sentiment-based research done that utilizes the Dodds et al. Hedonometer to some extent. In particular, they present the accuracy of several methods of topic detection including bag of words, classification, and deep learning.

Mitchell et al. presents a formal definition of happiness in the context of geolocated communication and friend networks.

Volkova et al. is a research group whose interests are strongly connected with mine. Topics involving gender, age, political, religious inference - and the resulting models are strongly correlated with the study I am doing.

3 Approach and Implementation

For this study, there will be no training set. Technically, since we are leveraging some tools which are already well trained (sentiment140)[6], we cannot call this approach purely "unsupervised". Additionally, the *stretch purpose* of this project is to develop an accurate model - in which words are considered features - to predict the happiness of users based off of their social media content. In the event that we reach this goal, either elements of a test set will have to be gathered in order to evaluate and improve our model, or we can test against our own correlation model by hiding the classification result from the prediction model.

Tools

In conducting this study, we plan on leveraging the Python programming language and utilizing multiple packages and libraries. Some important packages utilized in our preliminary research include

- Requests - http for humans
- Pickle - converts a Python object hierarchy int a byte stream, or visa versa. In this case, we use this package to store the sentiment hierarchy to a json file
- Pygooglechartapi - wrapper for the google chart api that allows us to easily create a chart and embed it in a web page as a png.
- Python-twitter - wrapper for the twitter api
- Sentiment140 - cloud based sentiment analysis

Dataset

To gather a population of users, we crawl About.me's user database to grab social media accounts from a large population. About.me serves as a centralized repository for user's social media accounts. Leveraging about.me's api, we can parse multiple social media accounts from a single user to garner as much data as possible.

The About.me api uses a REST interface. An example query structure and json return is provided in Appendix A.

Since we want to derive knowledge from as accurate data as possible, instead of using machine learning techniques to classify the religion/relationship status/etc. of a user based on twitter posts, we can obtain the ground truth by reading the relationship field from their Facebook. Since About.me has a broad and populated userbase, we can afford to be selective and only gather data from those users who have filled the relationship status field on any one of their social media accounts. Currently, however, I have limited my query size to 2,000 due to complications with About.me's random user retrieval[1]

Preliminary Findings

Preliminary research has been conducted in order to find an extensive dataset and successfully extract basic sentiment and knowledge from social media data. Easy analysis has been done as well as implementation of a basic visualization system. In particular, we can parse a users social media data, compile it into a uniform structure, provide the data to sentiment140 for analysis and present the results as an html webpage. An example result for individual users is provided below for three presidential candidates where the x axis is the post number and the y axis is given as the sentiment result on a per post basis. Functionality will be added for a post/time-based visualization, stream data, multi-user compilation, and legitimate happiness metric. [2]

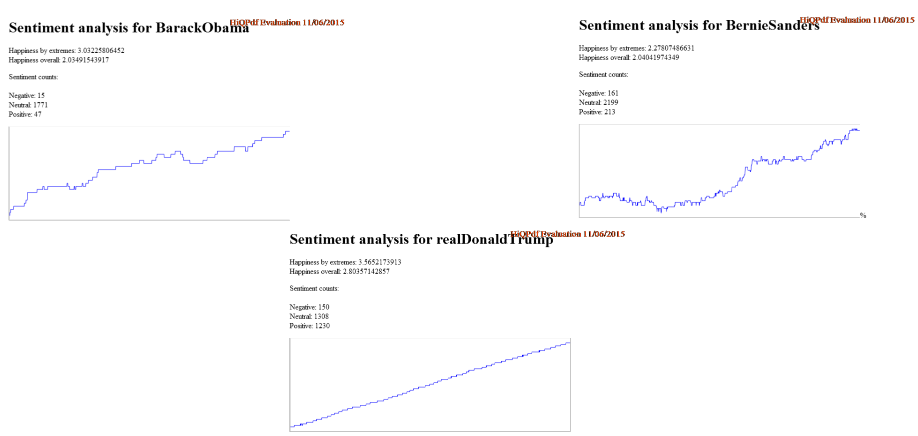

Figure 1: Presidential Social Media Sentiment[6]

References

[1] Cody, Reagan, Mitchell, Dodds, Danforth, *Climate change sentiment on Twitter: An unsolicited public opinion poll*, [arXiv:1505.03804].

[2] Daswani, Leike, *A Definition of Happiness for Reinforcement Learning Agents*, [arXiv:1505.04497].

[3] Dodds, Danforth, *hedonometer*, `http://hedonometer.org/index.html`.

[1] About.me's API seemingly does not sample users uniformly - there is some replacement that results in duplicate users being returned in the query result.

[2] See the figure citation for source.

[4] Frank, Mitchell, Dodds, Danforth, *Happiness and Patterns of Life: A Study of Geolocated Tweets*, [arXiv:1304.1296].

[5] Gilbert, Eric, Karahalios, Karrie. *Predicting Tie Strength With Social Media*, University of Illinois at Urbana-Champaign.

[6] Holtz, Chester, `https://github.com/Choltz95/Happiness_Companionship`.

[7] Preot-Peitro, Volkova, Lampos, Aletras, Bachrach, *Correlating Perceived Predicted Psycho-Demographic Attributes and Interests on Twitter*, `https://www.cs.jhu.edu/~svitlana/interests.html`, Penn, Hopkins, London, Microsoft.

[8] PBS, `http://www.pbs.org/thisemotionallife/topic/happiness/what-happiness`

Appendix

A.

Query:

```
https://api.about.me/api/v2/json/users/view/random/?
client_id=abcDqEFGrH0IzJKylmnBGoP2qr2ST7WvxySGX0SaBC4
&token=123.4567890987.65.4e321012345678f9123c66a4567890ae12fe3450d6789
&profile_number=1
```

Return:

```
{
  "status": 200,
  "total_count": 1,
  "result":
  [{
    "profile": "http://about.me/melissadipasquale",
    "user_name": "melissadipasquale",
    "first_name": "MELISSA",
    "last_name": " DI PASQUALE",
    "display_name": "MELISSA  DI PASQUALE",
    "header": "Photographer",
    "bio": "Melissa graduated from Humber College during the spring ...",
    "background": "http://about.me/.../melissadipasquale_1290058842_91.jpg",
    "mobile_background": "",
    "email_searchable": true,
    "email_public": true,
    "avatar": "",
    "tags": [],
    "img_base_url": "http://about.me/.../thumbnail",
    "thumbnail_291x187": "http://about.me/.../291x187/melissadipasquale.jpg",
    "thumbnail1": "http://about.me/.../803x408/melissadipasquale.jpg",
    "thumbnail2": "http://about.me/.../260x176/melissadipasquale.jpg",
    "is_fav": true
  }],
  "total_available": 1
}
&profile_number=1
```