Title: Understanding You through your Instagram – Transferring artificially generated image captions to personal description.

Social media today, for many, is the predominant way of interaction, and receiving and sharing of information with others. People are uploading their most personal, and intimate moments via photos, and videos, and it has become the go-to medium for communicating and documenting apsects of one's life. According to a survey by Global Web Index, in 2018, internet users were spending an average of 2 hours and 22 minutes per day on social networking and messaging platforms. This is an increase of 57% from 2012, when the average usage rate was at 1 hour and 30 minutes. One major social media platform is Instagram; it is the largest photo sharing site in the world. It’s estimated that 35 percent of American adults have an account with over a billion active users globally. Every post consists of photo(s) or video(s) along with some caption and keywords which documents a particular sentiment or event. Individuals lean towards posting contents that are generally associated with personal interests, values or specific industries such as fashion, fitness, travel, food, music, politics, etc. Thus, every post gives a purview of the poster. With its popularity along with the advance of digital technology and media production, average person can now produce high quality media at high frequencies. This means that there is an abundance of data points to analyze the characteristics of many users. The aim of this work is to build a computation model that would perform a mid-to-high level characterization of a person's life from only the Instagram posts.

There is a deep rich information in social media images and videos, but the data is high dimensional and complex which contains a lot about the person that a simple classification algorithm cannot output. And most of posts on social networks hardily contain detailed descriptions of the image. Widely used image recognition models like the convolutional neural networks (CNNs) are excellent in labeling images, yet it falls short in contextualizing and summarizing the image on its own. To address, the solution is to generate captioning an image using a combination of CNN and recurrent neural network with pre labeled images with captions. The caption may provide greater contextual summary of image, and I predict that having a robust and informative feature is sensible for characterizing individual posts. Most of the work in this project will focus on creating an intensive image labeling framework using pretrained CNNs in conjunction recurrent models. Moreover, complementary and similar to visual content analysis, I will attempt to extract relevant features from user generated captions of the post as well.

All output features will undergo domain specific fine tuning to ensure sensibility and accuracy of the label and the features are weighted by the output probability of each model and combined. This process will require some handpicked parameter tuning but it will be mainly based on the data. With these feature set, we can categorize user characteristics using unsupervised grouping or clustering models. The model will be analyzed to test the validity of generated feature set. Finally, we will create a data analysis tool for characterization of individual Instagram user and a web app for visualizing the analysis along with features which will summarize the characterization of individuals circle of followers and followings.

The caveat of is that user has 100 percent control of what they choose to share but selective about what they want to be exposed to. The content a person is exposed to. Given the ease of accessibility and the sheer volume of content, creates a huge diversification.

**Image Caption Model Training**

**Data Collection**

To obtain data, first I needed a data miner for Instagram. For this, I’ve used a combination of third-party API for Instagram data scraping, one was InstagramAPI and the other was InstagramScraper. These two APIs individually were not sufficient for me to accomplish my goals which was one, to get the following and follower information of my account and two, download images, captions, comments and number of likes for every single post that a user posted. Two address this, I wrote a python package called InstaScraper which combined the two APIs that allowed me to quickly download data from any user which were public or that I was following. This part was the most challenging and most time consuming

**Data Pre-Processing**

Now that I was able to get the

**Modeling**

The goal is to understand how I can apply data science to First, to understand what kind of images that are being shown, I used a [Convolutional Neural Network](http://cs231n.github.io/convolutional-networks/). Because my classification problem was extremely detailed & subjective, I needed an algorithm that could extract a large enough amount of features to detect a difference between the profiles I liked and disliked. A cNN was also built for image classification problems.

In a survey of 112 people, which were mostly my friends and acquaintances on Facebook, I asked whether if they were following a social media “influencer” 82% responded say yes. And I asked a follow-up question asking if they themselves had interest of becoming one in the future in which 43% said yes.

Over the years, businesses and brands have paid big sums of money to social medial platform for advertisements and the social media companies, internally, have invested huge resource in trying to understand what people are interested and how they behave as consumers. With massive amount of user data, and the ability to analyze and interpret, these companies can segment and target audiences more quickly and accurately than ever before.

In a survey of 112 people, which were mostly my friends and acquaintances on Facebook, I asked whether if they were following a social media “influencer” 82% responded say yes. And I asked a follow-up question asking if they themselves had interest of becoming one in the future in which 43% said yes.

Over the years, businesses and brands have paid big sums of money to social medial platform for advertisements and the social media companies, internally, have invested huge resource in trying to understand what people are interested and how they behave as consumers. With massive amount of user data, and the ability to analyze and interpret, these companies can segment and target audiences more quickly and accurately than ever before.

Apps like Instagram has fundamentally altered the concept of socializing as more people interact over the web than in person. Instagram generates astronomical amounts of content through its user base. Just like any popular media medium, the content has power to form people’s perception and opinions

In one way or the other, every single post that is shared or viewed on Instagram influences us.

of the. and anyone from anywhere can share and see their content