Find User Interests with Text and Image Analysis of Social Media Posts

Project Status Report 2

Jake Lasley  
 Computer Science  
 The University of Texas at El Paso  
 El Paso, Texas  
 jlasley@miners.utep.edu

Alan Motta  
 Computer Science  
 The University of Texas at El Paso  
 El Paso, Texas  
 amotta3@miners.utep.edu

ABSTRACT

Social media and social media applications are hotspots for collecting information about individuals. These websites and their component applications are an outlet for people to share their interests, details about themselves, and react to other individual’s posts. The nature of these websites brings us closer to the people and business that we know and love. We add our friends online, we follow the artists we like to listen to, and we check in to the places we love to go to. Furthermore, they are great places for users to express their interests, not only through words, but also through pictures. With the given information about user’s interests (including their follower’s interests), we then make recommendations to users for potential new interests.

**STATEMENT OF PROBLEM**

There is an extremely large volume of digital content uploaded to social media sites, much more than any one person could consume in a lifetime. Many social media users use these sites to find information about activities and events in their areas. With so much information to go through it can be hard to find some new interests. By analyzing the user’s posts we can provide suggestions for new interests.

**OBJECTIVES**

By comparing a set of terms from a user to a set of terms attributed to a topic, we can give a probabilistic estimate of that user’s interest in the topic. In order to do this, we need to capture text from the user. For this we turn to Instagram posts. The contents of the posts are available through the Instagram API, which can be accessed through a python wrapper. The text content of posts can be extracted and stored for performing the analysis.

There are some different approaches to profiling the users. While the users can be profiled based on the content of the post text alone, they can also be profiled based on their communities. Communities are defined as the set of local friends, people they work with, and people they are in the same hobby groups with. (K. Ikeda et al.). We would like to see if there are similarities between the interests of the users in a certain area, and differences between different areas.

PROGRESS OF PROJECT

So far, we have been able to retrieve the post-information of approximately 670 users in the El Paso area. Building the data sets has been difficult because the API for instagram is private, meaning it can only be interacted with through a user account, and the sampling rates are slow. We are continuing to build the data set for El Paso, and will then add 2 more cities, New York, New York, and Los Angeles, California. To determine the location of the users, we have been using the location declared in their bio, if they have a location declared. If users do not have their location declared, then we keep them out of the data set.

To calculate a probability of a user’s interest in a topic we also need sets of words associated with the topic. These sets are traditionally large, and the topics more broad, for example, ‘people 40-50 years old’. Depending on the topic, the data sets change in size.

NEXT GOALS

Our next goal is to start performing analysis on the data set. We plan on using Aikake’s Information Criteria, AIC, to evaluate the relevance of the topics to the users. In the AIC model, terms can have positive and negative relevance. Terms with positive relevance are associated with, and appear frequently in text discussing the topic, and terms associated with, and appear frequently around other topics have negative relevance. Using this, we can calculate a probability of interest in each topic for each user.

REFERENCES

[1] How to Find and Reach Your Target Audience on Instagram, https://sproutsocial.com/insights/instagram-target-audience/

[2] 6 WAYS TO CREATE A MORE ENGAGED INSTAGRAM AUDIENCE, https://www.businessinsider.com/instagram-demographics-2013-12

[3] Top Instagram Demographics That Matter to Social Media Marketers, https://blog.hootsuite.com/instagram-demographics/

[4] Twitter user profiling based on text and community mining for market analysis

www.elsevier.com/ locate/knosys

[5] Towards User Personality Profiling from Multiple Social Networks

Kseniya Buraya,1 Aleksandr Farseev,2 Andrey Filchenkov,3 Tat-Seng Chua4

[6] Twitter user profiling based on text and community mining for market analysis

Kazushi Ikeda, Gen Hattori, Chihiro Ono, Hideki, Asok, Teruo Higashino.

Conference Name:ACM Woodstock conference

Conference Short Name:WOODSTOCK’18

Conference Location:El Paso, Texas USA

ISBN:978-1-4503-0000-0/18/06

Year:2018

Date:June

Copyright Year:2018

Copyright Statement:rightsretained

DOI:10.1145/1234567890

RRH: F. Surname et al.

Price:$15.00