

CS 398 - Undergraduate Research Report

Shelby Ruettiger

I. INTRODUCTION

The Computational Population Biology Laboratory at University of Illinois at Chicago explores the intersection between biology and computer science. Projects include using computer science to analyze the movement and behaviors of social animals, dynamic network analysis as it relates to population biology, automatically identifying patterned animals from photographs, and estimating wildlife populations via social media. Specifically, Wildbook is a project that uses technology, wildlife research, and citizen science to understand global populations of patterned animals by detecting individuals in photographs uploaded to the Wildbook project. The project discussed in this report relates to Wildbook and the Animal Wildlife ESTimator using SOCIAL MEDIA (AWESOME), an initiative that mines images and data from social media platforms such as Bing and Flickr to estimate wildlife populations.

II. PROJECT

Photos posted on social media usually host user and photo data through application program interfaces (APIs). This project sought to obtain pertinent user and photo data via the Flickr API for downloaded images from Flickr of the Grevy's Zebra. A small application written in Python named the flickrPhotoStatRetriever was created to compile social media data for each of these images. Pertinent user data collected from each photo included

- date of a user's first posting on Flickr
- date of a user's first photo taken
- number of photos posted by user
- user location
- whether or not the user paid for a professional account

Photo data collected for each photo included

- photo location
- number of views
- number of comments
- number of favorites
- date posted
- date taken

Latitudes and longitudes of user and photo locations were obtained via the Google Maps API to allow for fine scale locations.

This data will be used by Masters students to understand biases in social media postings of patterned animals and its relevance to calculating animal populations via social media.

III. RESULTS

Of 2121 images, the application successfully pulled data for 2118 images. The data from three images was unable

Account Type	Count	Mean	Median	Top 75%	Std Dev
Pro	1294	109.37	19	107	222.91
Regular	674	84.29	3	43	225.88
All users	1968	100	13	86	224.15

TABLE I: Pro Accounts vs Regular Accounts vs All accounts

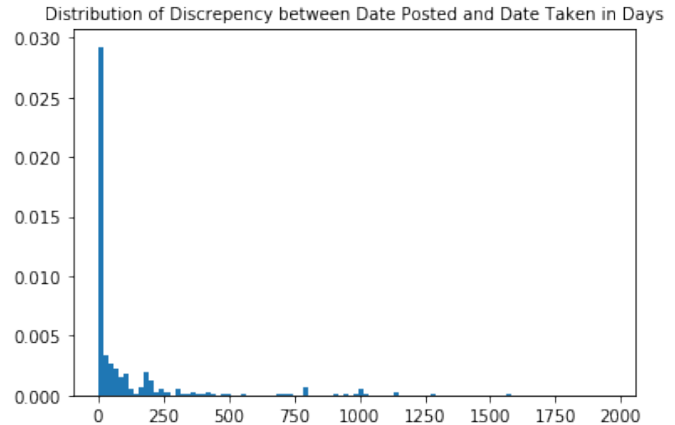


Fig. 1: Distribution of number of days between date taken and date posted for all users

to be obtained due to users deleting their accounts since the images were downloaded. One distribution of interest was the discrepancy of the date between the date taken and the date posted on Flickr for each photo. The median amount of days between the date an image was taken and date it was posted was 13 days and 75% were within 86 days. Additionally, a comparison of this date discrepancy was plotted between those users who had a professional Flickr account and those that had regular accounts (Figure 2). Table 1 shows descriptions of each dataset for comparison.

It is apparent from Table 1 that users with regular accounts post pictures to Flickr sooner than users with professional accounts on average.

Data on image locations and user locations was collected and plotted as well. Figures 3 and 4 show global locations of images and users, respectively.

IV. CONCLUSION

In conclusion, this undergraduate research project successfully mined data from social media platform Flickr in order to understand biases in social media posting by Masters students in the future. In an early analysis of a small section the data, it seems that regular accounts tend to post pictures sooner after the image is taken than those who pay for professional accounts. The application also successfully used the Google Maps API to return fine scale locations that were able to be plotted.

Distribution of Discrepancy between Date Posted and Date Taken in Days: Pros vs Others

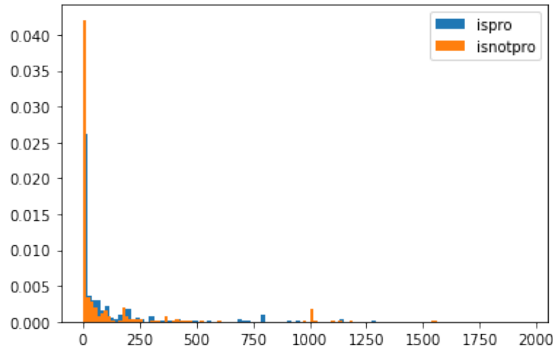


Fig. 2: Comparison of distribution of number of days between date taken and date posted - Pro vs Regular accounts

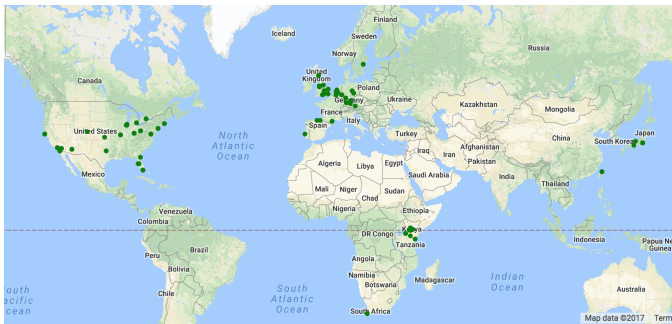


Fig. 3: Map of image locations

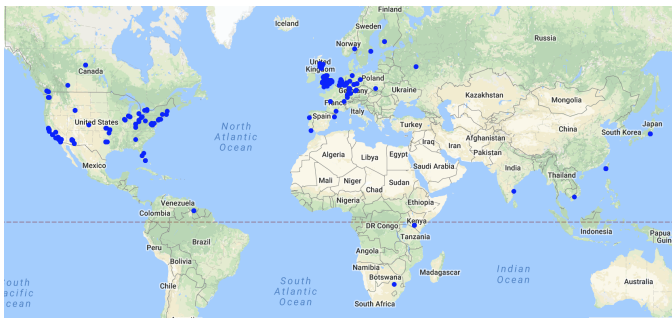


Fig. 4: Map of user locations

ACKNOWLEDGMENT

I would like to thank Dr. Tanya Berger-Wolf, Lorenzo Semeria, and Matteo Foglio for their support and guidance throughout the semester.