Social Media Image Analysis

Jiaming Zhao, Paine Leffler, & Eleanor Leffler

Problem

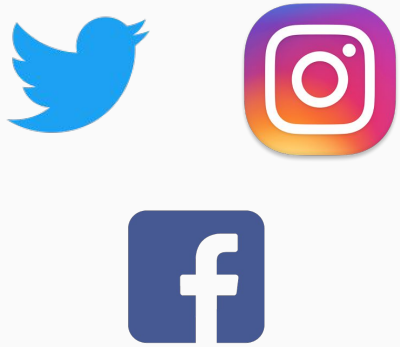
1. Less textual data available about people.

2. Excluding images excludes a large amount of data

3. Looking at hundreds of images from different social media sites takes time.

Goal

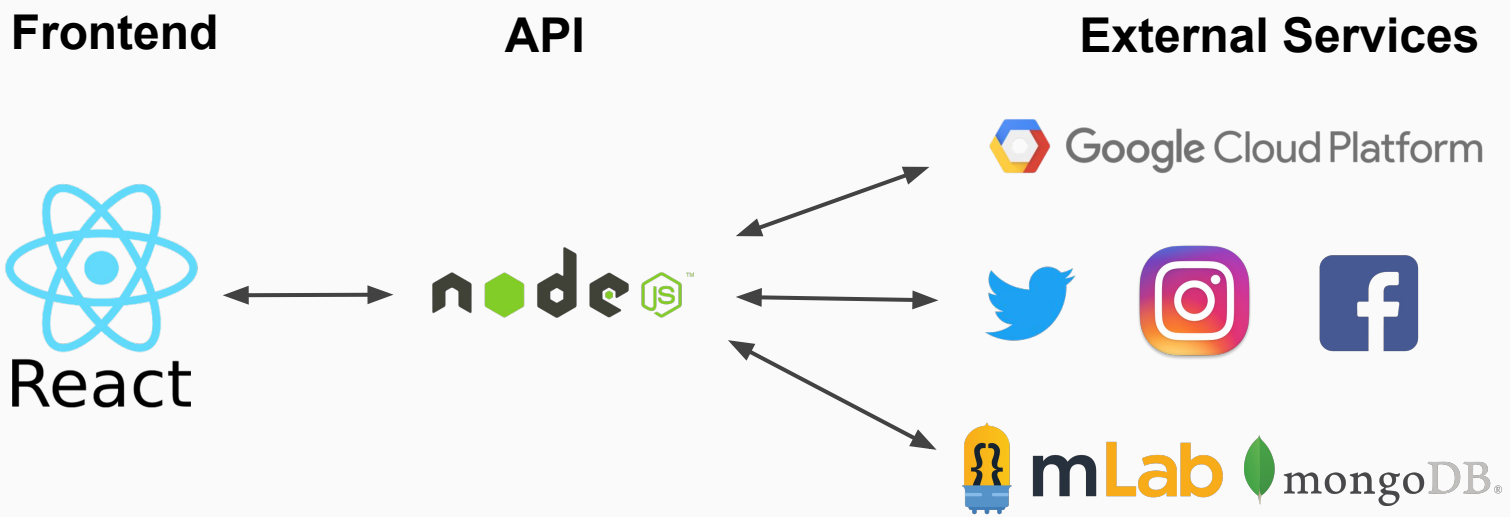
Images



Words & Numbers

Sports 63… **Politics** 12 … **Science** 10… **Technology** 79… **Outdoors** 52… **Dogs** 100…

Architecture



API Wrapper Routes

GET

Returns array of image URLs.

/twitter/images

/instagram/images

/facebook/images

POST

Search Mongo for URL. If not found, send image URL to Google and add results to MongoDB.

/results

Google Cloud Platform

GET

Returns labels and occurrence of specific user. Labels can be filtered by score, wordlist, and platform.

/labels

mongoDB.

Database

{

“Username”:

“Url”: *Direct Image URL*

“Platform”: *‘Twitter’, ‘Facebook’, ‘Instagram’*

“Results”: [{ }] *Google Vision Result Object*

“Date”: *Date Results were generated*

}

Twitter Response Data


```
"created_at": "Fri Jul 29 16:16:33 +0000 2016",
"id": 759060050589528065,
"id_str": "759060050589528065",
"text": "Wow\n          Such song https:\/\/t.co\/qWgRFrHvNa",
"truncated": false,
"entities": {
        "hashtags": [],
        "symbols": [],
        "user_mentions": [],
        "urls": [],
        "media": [{
                "id": 759060043283050496,
                "id_str": "759060043283050496",
                "indices": [27, 50],
                "media_url": "http:\/\/pbs.twimg.com\/media\/Coi5FnMUMAA5pJk.jpg",
                "media_url_https": "https:\/\/pbs.twimg.com\/media\/Coi5FnMUMAA5pJk.jpg",
                ...
```

Instagram Scraper Data

[{
  “id”: String - Post's ID from Instagram's DB,
  “caption”: String - Post's current caption,
  “shortcode”: String - Post's shortcode from Instagram's DB,
  “image”: String - URL for the post's image,
  “dimensions”: Object - Has properties 'height' and 'width',
  “likes”: Object - Has property 'count'
  “comments”: Object - Has property 'count',
  “owner”: Object - Has property 'id' that identify the owner of the post,
  “timestamp”: Number - Timestamp of when the post was created
}, {...}...]

Facebook Data

GET /search?q=JohnSmith&type=user

```
{ data:
   [ { name: ‘John Smith’, id: '402158663536002' },
     { name: ‘John Smith’, id: '10155395167713220' },
     { name: ‘John Smith’, id: '10215099113573489' },
     { name: ‘John Smith’, id: '1754786444593458' } ]
}
```

GET /page/

```
{
  "full_picture": "https://external.xx.fbcdn.net/...",
  "picture": "https://...",
  "id": "196118303833398_362048290573731"
}
```

Google Vision Response Data


```
"results": [
    {
      "faceAnnotations": [],
      "landmarkAnnotations": [],
      "logoAnnotations": [],
      "labelAnnotations": [
        {
          "mid": "/m/03qtwd",
          "locale": "",
          "description": "crowd",
          "score": 0.9742088913917542,
          "confidence": 0,
          "topicality": 0,
          "boundingPoly": null,
          "locations": [],
          "properties": []
        },

            ...

      "textAnnotations": [],
            "fullTextAnnotation": null,
            "safeSearchAnnotation": null,
            "imagePropertiesAnnotation": null,
            "cropHintsAnnotation": null,
            "webDetection": null,
            "error": null
        }
      ],
      "time": {
        "$date": "2017-11-10T00:21:09.706Z"
      }
}
```

Uninteresting Labels

Stored in MongoDB results, but not sent with response.

“product profession official photograph photography fun event snapshot room logo brand fiction graphics audio circle purple photo caption number line angle color white red yellow blue black screenshot text label girl woman man boy mammal animal material font area advertising advertisment . . .”

Interesting Labels

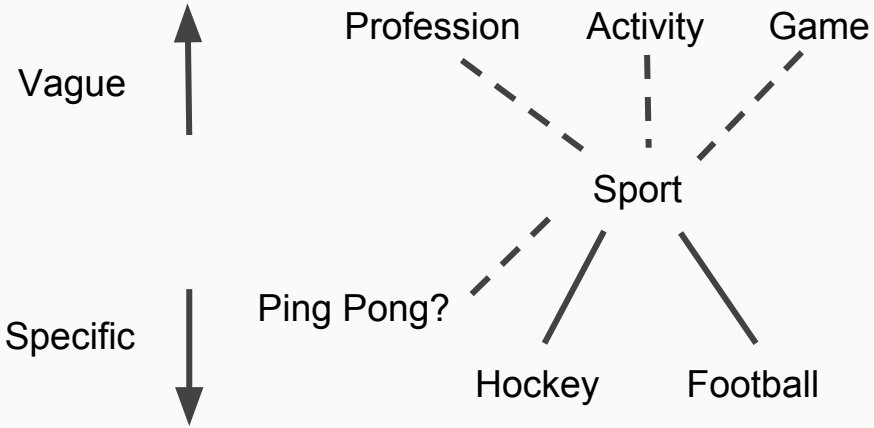
“Guitar, String Instrument Mountains, Lake, Kayak, Alcoholic Beverage, Football, Hockey, Coffee, Corgi, Motor Vehicle, Wilderness, Canyon, National Park, Cycling, … “

etc.

These distinguish people rather than giving high level attributes of the image.

Categorizing Labels

This requires comprehension of the language and can be subjective to opinions. Words often have multiple categories and it is difficult to put them in a category.



Demo

Instagram Issues

1. Restricted API Usage
2. Outdated Github Scraper repos / Outdated Forks
3. Scraper that worked… For a week…
4. Instagram Updates

Vision Issues

- Rate Limiting **requests/min**
- Label Condensing **ex: billiard table, pool, english billiards, snooker, cue sports**
- URL Restrictions

Good URL:

https://scontent-atl3-1.cdninstagram.com/t51.2885-…507577552_4145875980289835008_n.jpg

Bad URL:

https://instagram.fbna1-2.fna.fbcdn.net/t51.2885-...507577552_4145875980289835008_n.jpg

***Google**: "We cannot access the URL given"*

Vision Context Issues

Google Label: Desktop Wallpaper

Actually: Antelope Canyon, Page, AZ

Facebook Issues

- We did not have access to **/id/albums** for profile pictures
- Cant search on Facebook Username
- Facebook scrapers only provided basic profile information with the current profile picture and cover picture.

```
{ error:
  { message: '(#10) Application does not have permission for this action',
    type: 'OAuthException',
    code: 10,
    fbtrace_id: 'BI18yzyVgj9' } }
```

Real Use Cases

Recommendation Algorithm:

- Pages / Interests
- Friends / Public Figures
- Dating Apps
- Posts / Videos

Tools:

- Job Recruiting
- Team Selections
- Accessibility Plugins
- Branding/Marketing

Conclusions

- This Google Vision’s technology is great, but not perfect.
- Human feedback is needed to decide what labels are wanted.
- Because of retweets, image relevancy depends from person to person.
- Instagram pictures tend to be more personal.
- Word Classification is difficult.

Future Work

- Video/GIF analysis with frame by frame analysis.
- This combined with speech to text recognition would be an interesting way to learn about videos on social media or YouTube.

Thank You!

Questions?