

ANDROID APP RECOMMENDATION SYSTEM

Siddharth Jain

2018101038



Introduction

In this project I implemented a content based android app recommendation system (Implemented a "similar apps" feature of Google Play Store).

Datasets

I used two datasets :

1. Google apps dataset : Consisted of 10k app entries. This dataset had these columns : App Name, Category, Rating, Reviews, Size, Type.
 2. Google reviews dataset : Consisted of 63k app reviews, had reviews of 1k (way less than 10k) apps. This dataset had columns : App, Translated Review.
-

-
3. Created a new dataset from reviews dataset which was : concatenated reviews of apps, (one entry per app).

MODEL/ALGORITHM

The app similar to the given app must be of the same category as well.

To find similar apps to the app given by the user I used similarity_measure which was similarity in App name + similarity in App reviews.

Similarity is Cosine similarity of vectors, vectorized by Tfidf Vectorizer.

After sorting (descending) the similarity scores of all apps with the given app, I divided them into different buckets considering their similarity_measure. The max value of similarity_measure is 2, buckets were:

- 1.5 to 2.0
- 1.0 to 1.5
- 0.6 to 1.0
- 0.4 to 0.6
- 0.3 to 0.4
- 0.25 to 0.3
- 0.2 to 0.25
- 0.15 to 0.2
- 0.1 to 0.15
- 0.075 to 0.1

The threshold on similarity_measure was 0.075.

While recommending apps, apps in each bucket were sorted according to, their ratings, and the apps are recommended bucket wise, i.e., firstly the apps of bucket 1 (sorted wrt ratings) are recommended, then second bucket and so on.

Similar Apps for :

- Flipkart (category - SHOPPING) are:
 - Myntra Online Shopping App 4.3 0.505
 - Jabong Online Shopping App 4.1 0.505
 - GearBest Online Shopping 4.5 0.337
 - The Coupons App 4.5 0.322
 - Shopee - No. 1 Online Shopping 4.3 0.364

- Dropbox (category - PRODUCTIVITY) are :
 - Amazon Drive 4.4 0.251
 - G Cloud Backup 4.6 0.179
 - AndroZip™ FREE File Manager 4.2 0.151
 - ColorNote Notepad Notes 4.6 0.11
 - Adobe Acrobat Reader 4.3 0.129

- Foodpanda (category - FOOD_AND_DRINK) are:
 - Uber Eats: Local Food Delivery 4.2 0.51
 - BeyondMenu Food Delivery 4.7 0.303
 - Talabat: Food Delivery 4.5 0.303
 - DoorDash - Food Delivery 4.5 0.303
 - Grubhub: Food Delivery 4.5 0.303

-
- Google pay (category - FINANCE) are:
 - CB Pay 4.5 0.452
 - GCash - Buy Load, Pay Bills, Send Money 4.0 0.339
 - TNEB Quick Pay Easy 4.5 0.27
 - Bank of America Mobile Banking 4.4 0.254
 - Amex Mobile 4.3 0.254

 - Expedia (category - TRAVEL_AND_LOCAL) are:
 - Booking.com Travel Deals 4.7 0.654
 - Hotels.com: Book Hotel Rooms & Find Vacation Deals 4.5 0.617
 - CheapTickets – Hotels, Flights & Travel Deals 4.4 0.638
 - Hotwire Hotel & Car Rental App 4.3 0.726
 - Hotels Combined - Cheap deals 4.1 0.628

 - BBC News (category - NEWS_AND_MAGAZINES)
 - Fox News – Breaking News, Live Video & News Alerts 4.5 0.65
 - Fast News 4.4 0.609
 - BBC Persian | BBC Farsi News 4.3 0.79
 - AP Mobile - Breaking News 4.5 0.508
 - ARY NEWS 4.4 0.408

 - Hinge: Dating & Relationships (category - DATING):
 - BBW Dating & Curvy Singles Chat- LargeFriends 4.4 0.308
 - CMB Free Dating App 4.0 0.379
 - BeWild Free Dating & Chat App 3.4 0.349
 - Free Dating Hook Up Messenger 3.3 0.385
 - Healthy Relationships Guide nan 0.362
-

-
- Cisco Webex Meetings (category - BUSINESS):
 - Cisco Webex Teams 4.2 0.853
 - ZOOM Cloud Meetings 4.4 0.319
 - join.me - Simple Meetings 4.0 0.298
 - Facebook Pages Manager 4.0 0.109
 - HipChat - Chat Built for Teams 3.8 0.142

Conclusion:

The recommendations seem to be good enough but due to the lack of reviews (issue with the “reviews” dataset), only apps whose name starts with A-H had reviews in the review database and not others. Due to this, the recommendations are not perfect.