2/8/2017

Mobile Game Analyzer

CS838 Project Report – stage 1

GAURAV MISHRA [gmishra2@wisc.edu]
OM JADHAV [ojadhav@wisc.edu]
PALLAVI MAHESHWARA KAKUNJE [kakunje@wisc.edu]

1. Introduction

Over the past few years, apps have become an essential part of the mobile user experience. With millions of apps now available and more being rolled out every day, marketers and app developers might find it useful to know what kind of apps are more popular among users, to continue to add functionality to apps, to stand out in an increasingly competitive marketplace.

In this project, we target to analyze game apps from two giant apps stores: Google PlayStore and Apple's AppStore and answer few of the queries which might be useful. We also intend to learn documents containing game reviews and extract some useful information.

2. Questions we answer

1. Determine whether the application is more popular in PlayStore or AppStore by comparing the user ratings and download count.
2. Average app rating considering ratings from both the sources.
3. Guess the download count of apps in AppStore (which is not currently available) by analyzing the ratio of user rating and download count in PlayStore.
4. Total download count considering information from both the sources.
5. Subcategory (Ex: arcade, adventure, trivia etc.) which has more apps.

3. Data Sources

We extracted our data sets from the following two sources.

Data set 1: Google PlayStore

Data set 2: Apple AppStore

This project focuses only on game apps. Our data sets include a good mix of game apps from all the available subcategories: Action, Adventure, Arcade, Board, Card, Casino, Casual, Educational, Music, Puzzle, Racing, Role Playing, Simulation, Sports, Strategy, Trivia and Word.

We also extracted 525 documents from the website http://toucharcade.com/. It contains text describing game reviews.

4. How did we extract Structured Data

1. We identified the attributes for extraction. (title, developer, version, number of ratings, final rating, updated date, reviews of customers, description, price, language of the game)
2. Used Scrapy tool to scrape the data from the app store of Apple and Play store of Google
    a. We started with the main site which has list of all the apps,
    b. We followed links to reach to every application page and gathered above attributes in json format.
    c. We have gathered data of around 3700 games from App store and 4000 games from Play store

d. We have also gathered top five reviews from app store separately of around 1000 games.

5.What we extract from the text documents

1. Identify the category of the game (i.e. Action, Adventure, Board, Puzzle, Card, Educational, Music, etc.)
2. Predict if the review is a positive or negative (sentiment of the review).

6.Open-source tools used

6.1.Scrapy

We used scrapy to extract all the data sets and documents.

It is a free and open source web crawling framework, written in Python. It provides a fast and simple way to extract data from websites. Scrapy project architecture is built around 'spiders', which are self-contained crawlers that can take set of customized instructions to get the desired fields from the page in a specific format.

6.2.Jupyter Notebook

We used this to collaborate and develop/test our primary code.

The Jupyter Notebook is a web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text. In the recent years, Jupyter Notebook has become one of the important tools for data scientists.