E-Health Methods and Applications: Report Part 1

Contents

Retrieve Google-Playstore.csv Dataset from Kaggle

We decided to work on a Google-PlayStore database provided by a Kaggle repository. It was the most detailed one. Moreover, the primary key of the database is the appId. This attribute is used also with the google-play-scraper library.

Filtering the Applications by Relevant Educational Features

We chose those parameters based on our assumptions of what a "real" serious game should be. We considered the feedbacks provided by the users of the app as the most reliable definition of "how good is an application".

Thus, the filter selects an application if:

- It belongs to one of the educational categories ("Education", "Educational", "Family", "Learn", "4-year-old kids" and "4-year-old")
- Its rating is greater or equal than 4
- Its rating counts are greater than 50 000

We chose the categories that are semantically related to education and children. The minimum rating of 4.0 represents most of the applications, since the average of good applications is far above this value.

Indeed, if a game has a rating of 5/5 but only has a few ratings, the rating may not be very representative of the quality of the application. We agreed on rating counts greater than 50 000 according to our data source and the number of applications in the final dataset.

We found that other attributes were not as reliable as we thought they could be. For example, the number of installs sometimes does not reflect how famous or not is an app. Some apps could be installed by default or as a dependency.

Filtering the Games

We noticed that each application has a specific sub-category. It always starts with the string "*GAME_*" in case of a game. We checked the subcategory of each educational app with google-play-scraper in order to remove all those apps that were not games.

Furthermore, since our data source is an old dump of the real Google-PlayStore. Some apps could not be found with google-play-scraper because they were deleted or renamed. We chose to select and remove them from our dataset.

Enriching the Dataset with Descriptions and Reviews

Our dataset presents just the educational games with valuable features. Nevertheless, some important attributes are missing. Thus, we have implemented a first function that enriches our database with the description and the number of reviews of each educational game, using google-play-scraper.

We think that the description can be very useful to understand the purpose of a game and to judge how serious it is. Moreover, the description will be used in the next steps to extract other important parameters from the games. The number of reviews can be used in conjunction with the rating counts to analyze the trustworthiness of the rating of an application.

Using Natural Language Processing to Identify Learning Category and Age Range

We wanted to classify the different games according to their learning categories and age ranges. Since these parameters could not be found in the initial dataset nor with google-play-scraper, we decided to use Natural Language Processing. For each parameter, we first defined specific categories: ["science", "counting", "language", "creativity", "shape", "food", "music" and "sport"] for the learning categories and ["babies", "children", "adolescents", adults"] for the age ranges.

These categories have been selected to precisely separate the games. Then, we wrote a list of associated keywords by looking at synonyms or semantically related words for each one of these categories. From those lists of keywords, we wrote a function that counts the maximum number of keywords found in the applications' descriptions for each category, in order to assign the closest category to all of them

Export the Final Dataset

The original data source has been shrunk and enriched with new information. The new dataset was exported into a new csv file.

Benchmark

We chose 120 applications from the Google-PlayStore:

- 40 games judged as serious by the members of the group (human beings)
- 40 random misleading non-serious games that may deceive our filter
- 40 random applications.

This set of application was used to benchmark the preciseness of our algorithm. The evaluated statistical parameters are:

- *Accuracy*: the ability to distinguish between a serious game and non-serious one or any other application
- *Sensitivity*: the ability to recognize a serious game
- *Specificity*: the ability to recognize a non-serious game

We developed another parameter to test the accuracy, named *Effective Accuracy*. This parameter measures the ability to focus on serious games. The algorithm is penalized anytime that a mistake is made:

- Serious game not present
- Non-serious game present
- Random application present

Benchmark Results

Accuracy: 92.5 %

Sensitivity: 77.5 %

Specificity: 100.0 %

Relative Accuracy: 85.0 %