

Amazon cell phone evaluation platform

Marion Kramer

marion.kramer@epfl.ch

Alexandre Rassinoux

alexandre.rassinoux@epfl.ch

Kristina Satara

kristina.satara@epfl.ch

Abstract

Reviews help when making decision which product to buy, as they give possible customers personal experience and more product details. Since customers usually invest some time to studying the reviews and ratings before buying specific products from Amazon, we decided to create evaluation platform which would help the users decide about purchasing the product by giving them insight into downsides and advantages of various brands based on the reviews. The aim of this project is to analyze reviews of different brand of Amazon cell phone products and based on the reviews to decide which of the features are important for specific brands of cell phones. For this we use natural language processing techniques, such as tf-idf matrix, ... and Latent Dirichlet Allocation.

1 Introduction

Sentiment analysis, also known as opinion mining or emotion AI, is important part of natural language processing. The subject of the study is the attitude or the emotion behind certain text. The purpose of sentiment analysis is to determine whether text content is neutral, positive or negative towards some topic.

Data used for this project is collected from Amazon, during the period of few years. Most of the data is collected between 2012 and 2014, as shown in figure 2. Each of the reviews has also ratings that are used for ground truth. The ratings are based on 5-star system, where 5 is highest and 1 is lowest rating. Distribution of the ratings is J-distribution with highest number of 5-star reviews, and it is shown in figure 1. After purchasing some of the products customers are asked to give their review and rating, which are further subject to sentiment analysis in our project.

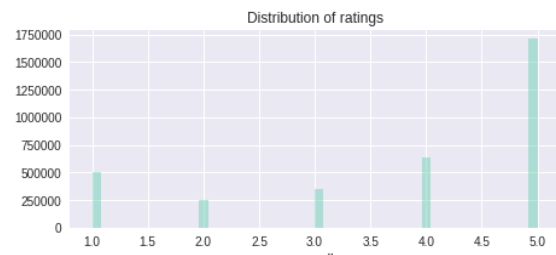


Figure 1: Distribution of the ratings in reviews

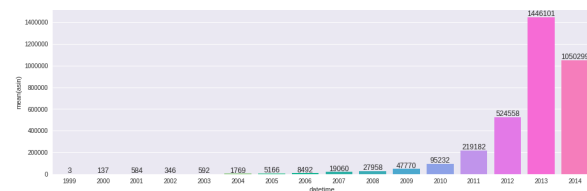


Figure 2: Number of reviews during time

2 Related work

Natural language processing is interesting and popular area of research, thus there are many papers regarding this topic. As Amazon is widely used platform, it has gained much reviews in the past few years. We focused on the cell phone reviews, and we found papers and which gave different approaches and conclusions. GIVE REFERENCES AND EXAMPLES.. In this paper they took this approach, etc etc.

3 Data Collection

We used dataset provided by Amazon for this project. Dataset consists of two json files - one containing the reviews, and another providing metadata about the products. Given below are the fields provided in the datasets:

- Reviewer's ID
- ID of the product reviewed
- Name of the reviewer

- Text of the review
- Rating given
- Summary of the review
- Time of the review
- Price in euros
- Url of the product image
- Brand name
- Category of the product

We merged these two files for our further data analysis. Also we transformed the data in order to improve the speed of loading and processing this dataset in our python notebook. As this dataset was incomplete, we decided to enrich it with additional dataset from Kaggle platform. This is public dataset from GSMArena and it contains more than 8000 phones specifications.

4 Dataset Description with Summary Statistics

TODO: Explain step by step everything from the notebook. Add some more pictures.

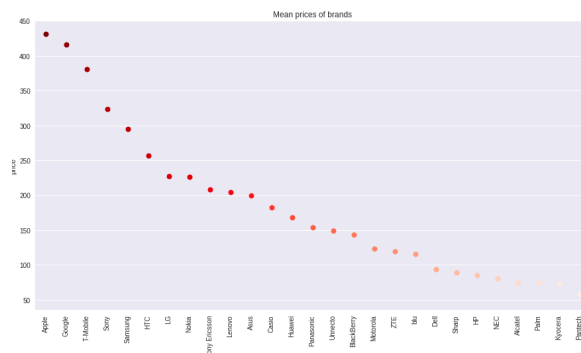


Figure 3: Mean prices per brand

5 Methods

Methods with math and description of main algorithm Explain ft-idf, LDA, and our algorithm.

6 Results and Findings

Elaborate the results about the interesting words into more details.

7 Conclusion

This platform uses reviews and extracts meaningful information from it, and then evaluates certain cell phone brand based on these reviews. We use sentiment analysis to classify each of the reviews (and it's parts) as good or bad. Based on this we describe and evaluate specific brands.

References

- Alfred V. Aho and Jeffrey D. Ullman. 1972. *The Theory of Parsing, Translation and Compiling*, volume 1. Prentice-Hall, Englewood Cliffs, NJ.
- American Psychological Association. 1983. *Publications Manual*. American Psychological Association, Washington, DC.
- Association for Computing Machinery. 1983. *Computing Reviews*, 24(11):503–512.
- Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. 1981. Alternation. *Journal of the Association for Computing Machinery*, 28(1):114–133.
- Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK.