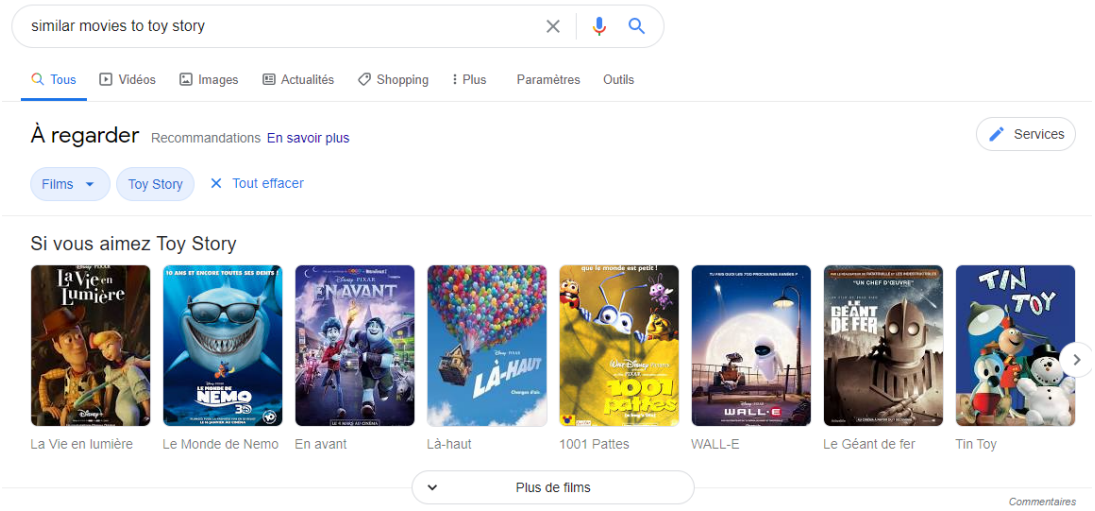
Movie recommendation system

Machine Learning - MSc DMDS - EM Lyon

Year 2020-2021

1 Introduction

Have you ever been on an online streaming platform like Netflix, Amazon Prime, Voot? You watched a movie and after some time, that platform started recommending you different movies and TV shows. But how could that movie streaming platform suggest appealing content ? They use a Recommendation System. This system is capable of



learning your watching patterns and provides you with relevant suggestions. Today, in the fourth industrial revolution where Artificial Intelligence and other technologies are dominating the market, you must have come across a recommendation system in your everyday life for sure. In this Machine Learning Project, you will build your own recommendation system.

2 Objectives / Tasks

The objective of the project is to determine a selection of 3 movies to recommend for a particular user. For this study, you are proposed to use this dataset:

https://drive.google.com/file/d/1Dn1BZD3YxgBQJSIjbfNnmCFlDW2jdQGD/view

This contains two files: the first file, contains informations about the movies and the second file contains informations about the user ratings.

The first task of this project is to explore how you can use that dataset to make the recommendations. This is called the pre-processing step. Here are a few hints:

- The films are classified with genres, create a one-hot encoding to create a matrix that comprises of corresponding genres for each of the films. This will bring you to a table that would look like:

```
movieId       title       Action Adventure Animation ...
      1   Toy Story            0         1         1 ...
```

- There are actually two problems: the similarity between users and the similarity between movies.
- A minimum number of review can be provided to discard movies for which only few data is available.
- Have a look at `cosine_similarity` from Scikit learn and what cosine similarity is.
- You might want to simplify the problem of user rating by just two values (like or not like - 0 or 1)

The second task of this project is to determine a training technique to be used for the machine learning. Ideally, the value predicted would be a movie type. Then you would get the best movies of that type that the user has not seen for now. It could also be directly a movie. This process will definitely involve classification or clustering.

The third task will be to train a machine learning model, which would analyze the dataset and predict, according to information of a particular user, one movie or a list of movies that would be recommended to the user. You will have to prove the validity of your model.

The fourth task is to prepare a presentation of your results. This implies making visuals to explain your results. You will have 15 to 20 minutes to show the result of your work.

3 Grading

The grade of your project will be determined by the teacher after your presentation. There will be no individual grade for members of the same team, unless there are clear disparity in the effort that was put. The four different task will each represent 25% of the final grade.