

POLITECNICO
MILANO 1863

NUMERICAL ANALYSIS FOR MACHINE
LEARNING
ACADEMIC YEAR 2021 - 2022

movieRecommendation

Sofia MARTELLOZZO Matteo NUNZIANTE

Professor

Edie MIGLIO

Contents

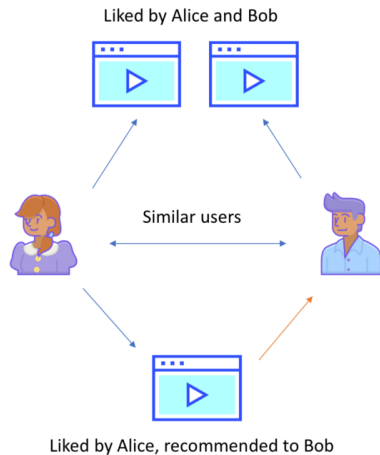
| | | |
|---|-----------------------|---|
| 1 | Introduction | 2 |
| 2 | Objectives | 4 |
| 3 | Dataset description | 4 |
| 4 | Algorithm description | 4 |
| 5 | Results | 4 |

1 Introduction

With the advent of the internet today, we are witnessing an enormous information overload. This exponential growth in data results in difficulty organizing and analyzing this basic information but opens up new avenues on the paths of knowledge. The question is no longer to have the information but to find the relevant information simultaneously; from there, recommendation systems were born.

Recommender System is a system that seeks to predict or filter preferences according to the user's choices. Recommender systems are utilized in a variety of areas including movies, music, news, books, research articles, search queries, social tags, and products in general. Recommender systems produce a list of recommendations in any of the two ways :

- **Collaborative filtering:** Collaborative filtering approaches build a model from the user's past behavior as well as similar decisions made by other users. This model is then used to predict ratings for items that users may have an interest in.

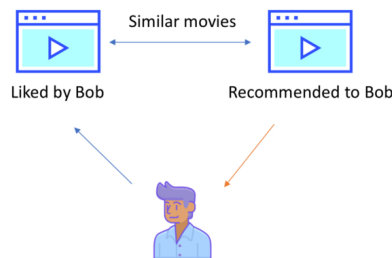


The advantages of this approach are:

- Domain knowledge not required: The system does not required a domain knowledge because is based only on item ratings.
- Serendipity: The model can help users discover new interests. In isolation, the ML system may not know the user is interested in a given item, but the model might still recommend it because similar users are interested in that item.

The limitations of this approach are:

- Cold start problem: The prediction of the model for a given (user, item) pair is the dot product of the corresponding embeddings. So, if an item is not seen during training, the system can't create an embedding for it and can't query the model with this item.
 - Sparsity: The system may find problems on predicting evaluation because of a situation in which the users evaluate a little of the total number of items available in a dataset. This creates a sparse matrix with a high number of missing values.
- **Content-based filtering:** Content-based filtering approaches uses a series of discrete characteristics of an item in order to recommend additional items with similar properties. Content-based filtering methods are totally based on a description of the item and a profile of the user's preferences. It recommends items based on the user's past preferences.



The advantages of this approach are:

- User autonomy: The model doesn't need any data about other users, since the recommendations are specific to this user. This makes it easier to scale to a large number of users. The model can capture the specific interests of a user, and can recommend niche items that very few other users are interested in.
- Immediate consideration of a new item: The model does not need the evaluation of all movie by a user, because it can be recommended without being evaluated.

The limitations of this approach are:

- Content too specific: The model can only make recommendations based on existing interests of the user. In other words, the model has limited ability to expand on the users' existing interests.

- Big scale information: Since the feature representation of the items are hand-engineered to some extent, this technique requires a lot of domain knowledge. The items must be enough detailed described and a user must evaluate several items before the system can interpret its preferences.

It is also possible to combine these two class of recommendation in order to overcome some limitations they faced. This type of approach is called **Hybrid Recommendation**.

In this project the items that has been avaluated are movies, and their score are the rating that the users gave them, is supposed after they whatched it.

2 Objectives

3 Dataset description

4 Algorithm description

5 Results