

Programming Assignment 2

Recommender Systems

Anísio Mendes Lacerda
April/2017

1 Introduction

Recommender systems are computational tools that help user to deal with the huge amount of information available. In this assignment, you are asked to implement different strategies to recommend items to users. The dataset used here refers to a recommender system that suggest scientific articles to users.

2 Files included in this exercise

- **pa2.pdf**: This file
- **dataset.tgz**: Dataset to be used in this assignment.

3 Recommender Systems Strategies

Please refer to [2] to understand the most common recommendation strategies proposed in the literature.

4 Deliverables

In this work, you need to implement five strategies for recommending scientific articles:

- *Random*: suggest a random article.
- *Most popular*: suggest the most popular article in the dataset.
- *Content-based*: use the abstract of each article to recommend articles to users. You are free to use any content-based method that you chose.
- *Collaborative filtering* (based on matrix factorization): use the userXitem matrix to suggest articles to users. Here, you are free to use any collaborative filtering strategy based on matrix factorization. Please refer to [3] to understand matrix factorization strategies.
- *Hybrid*: Combine the previous recommendations (i.e., based on content and collaborative information) to suggest articles to users.

IMPORTANT CONSIDERATIONS:

- Consider that you are recommending a list of articles to users, each recommended list must contain N articles. Vary the value of N as 1, 5, 10, 20.
- To evaluate your recommendation strategies use the metric known as nDCG [1].
- Report the averaged nDCG values over users and folds.
- It is your responsibility to split the dataset into $k = 5$ folds, i.e., you must use the k -fold cross validation methodology to compare the implemented strategies.
- You are free to use optimization libraries when implementing the matrix factorization strategy (e.g., `scipy.optimize`). Please, if you opt to use any library for optimization, point out which library you used and why.

References

- [1] nDCG. https://en.wikipedia.org/wiki/Discounted_cumulative_gain.
- [2] Jesús Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. Recommender systems survey. *Knowledge-based systems*, 46:109–132, 2013.
- [3] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8), 2009.