Tutorial Information Retrieval

Universität Innsbruck - Department of Computer Science

E. Zangerle

DBIS DATABASES AND INFORMATION SYSTEMS

2019-05-29

Exercise Sheet 8

Exercise 1 (RecSys Challenge, First Steps) [7 Points]

In the following weeks, we are going to participate in the RecSys Challenge, a challenge that is hosted by the yearly ACM Recommender Systems Conference. This year, the data, the tasks, and the evaluation platform are provided by trivago and the challenge is situated in the field of hotel booking platforms. The goal of this challenge is, in short, to predict hotels the user clicked on (so-called "click-outs") within a browsing session on the trivago platform and to derive a set of recommendations from these predictions.

In this first exercise, our goal is to get a deeper understanding of the task, the data and to set up our tooling for the challenge.

a) *1 Point* Please get accustomed to the task of the challenge. Go through the documents provided and the description of the task.

b) *3 Points* You should start with thoroughly analyzing the data you are provided with to get a first impression on how the data is distributed, what kind of information is contained and how sparse this information really is. Please note that in contrast to the MovieLens dataset, we are now dealing with full sequences of user interactions within a single session. Please perform an exploratory analysis of the dataset and already think about which information might be useful for the recommendation task.

c) *1 Point* The recommendation task will be evaluated by using the Mean Reciprocal Rank (MRR) metric. Implement the evaluation of the MRR (and possibly, further measures for analyzing the recommender system's behavior). You can find an example in the challenge's dataset description.

d) *1 Point* Please implement your own baseline approach - come up with a very basic approach to computing recommendations based on the data provided to make sure that your pipeline (reading in files, computing recommendations based on the training data and performing the evaluation based on the test data) works as expected.

e) *1 Point* To prepare your tooling for future submissions and hence, direct comparisons with other contestants, please make sure to create a submission file that fulfills the requirements for the submission files. You can also find further information on the Challenge's github repository. Please make a first test submission on the platform using your naive approach and make sure to be able to compute results for the provided baseline algorithm.

Exercise 2 (MRR Controversies) [3 Points]

Please read the following paper by Norbert Fuhr, in which he raises a few important points on MRR and further metrics in the information retrieval field. This should give you a critical perspective on evaluation setups.

Norbert Fuhr. Some Common Mistakes In IR Evaluation, And How They Can Be Avoided. *SIGIR Forum*, 51(3):32–41, 2018. ISSN 0163-5840. doi: 10.1145/3190580.3190586. URL http://sigir.org/wp-content/uploads/2018/01/p032.pdf

Answer the following questions:

a) *1 Point* Why is MRR (or ERR) not a good choice for assessing ranked lists, according to Fuhr?

b) *1 Point* What is the problem with MAP?

c) *1 Point* Why is it not advisable to use simple holdout evaluation?

Important: Submit your solution to OLAT and mark your solved exercises with the provided checkboxes. The deadline ends at 23:59 on the day before the discussion.