Evaluating Recommender Systems: Survey and Framework

Abstract

‘The comprehensive evaluation of the performance of a recommender system is a complex endeavor: many facets need to be considered in configuring an adequate and effective evaluation setting. Such facets include, for instance, defining the specific goals of the evaluation, choosing an evaluation method, underlying data, and suitable evaluation metrics. In this article, we consolidate and systematically organize this dispersed knowledge on recommender systems evaluation. We introduce the Framework for Evaluating Recommender systems (FEVR), which we derive from the discourse on recommender systems evaluation. In FEVR, we categorize the evaluation space of recommender systems evaluation. We postulate that the comprehensive evaluation of a recommender system frequently requires considering multiple facets and perspectives in the evaluation. The FEVR framework provides a structured foundation to adopt adequate evaluation configurations that encompass this required multi-facetedness and provides the basis to advance in the field. We outline and discuss the challenges of a comprehensive evaluation of recommender systems and provide an outlook on what we need to embrace and do to move forward as a research community.’

Link: https://dl.acm.org/doi/10.1145/3556536

The Netflix dataset is an ideal choice for our project as it aligns well with the FEVR framework introduced in the paper "Evaluating Recommender Systems: Survey and Framework." This dataset provides high-quality, well-documented data, essential for reliable evaluation. Its extensive historical ratings allow for robust offline evaluations of various recommendation algorithms, making it suitable for implementing and testing the techniques described in the paper.

Additionally, the Netflix dataset supports the calculation of key evaluation metrics, such as RMSE and MAE, which are crucial for assessing predictive accuracy and conducting user-centric evaluations. The dataset's structure also facilitates testing our hypotheses on the prediction accuracy of SVD-based models compared to item-based Pearson correlation methods. Overall, the Netflix dataset ensures a comprehensive evaluation of the recommender system's performance as advocated by the FEVR framework.

Link: https://www.kaggle.com/datasets/rishitjavia/netflix-movie-rating-dataset

Musaed owied – 2140988 Elyas Sindi – 2143512 Nawaf fayez alrehaili -2142609