Multicriteria Evaluation for Top-k and Sequence-based Recommender Systems

Diego Michele Monti

Summary

Context The need of effectively identifying relevant items from a potentially overwhelming catalog has led to the creation of automatic approaches for supporting users of online platforms in such a time-consuming task. Recommender systems are software tools and algorithms designed to suggest items to users according to their preferences. The traditional goal of the recommendation problem is to create a ranked list of suggestions for each user. However, a more novel paradigm is represented by algorithms capable of exploiting the temporal dimension of the available ratings, known as sequence-based recommender systems. Differently from the field of information retrieval, performing an offline experiment for comparing multiple recommendation approaches is a challenging task, as the ground truth is represented by the subjective preferences of the users collected before the introduction of the system under evaluation. Nevertheless, it represents a powerful tool for selecting the most promising approaches to be further tested in subsequent trails.

Goal In this dissertation, we study how to perform the offline evaluation of a generic recommender system by exploiting a multicriteria approach that relies on a set of heterogeneous metrics. We provide an answer to the following research questions: what is the current state-of-the-art regarding multicriteria recommender systems and how their are evaluated in literature; how can a multicriteria evaluation approach be exploited for comparing different sequence-based recommender systems; what is the most suitable protocol for performing an offline evaluation of a top-k recommender system; and to what extent the structure of a rating dataset can influence the results of an offline evaluation.

Method To answer the first question, we investigate the topic of multicriteria recommender systems with a systematic literature review. Regarding the second question, we propose an evaluation framework called Sequeval designed to compare in a replicable way different sequence-based recommenders considering eight dimensions. With respect to the third question, we introduce RecLab, an evaluation toolkit designed following a distributed approach to overcome some

limitations of the currently available evaluation solutions. Finally, for addressing the fourth question, we describe a method based on data visualization to explore the structure of a rating dataset and we report on an algorithm capable of generating alternative versions of an existing collection of ratings.

Results From the systematic literature review, we observed that is not possible to directly compare the results obtained with different multicriteria recommendation approaches due to the extreme variability in the reported experimental protocols. We exploited Sequeval and RecLab to successfully conduct some experimental campaigns involving different recommenders and datasets. The availability of multicriteria metrics enables the experimenter to obtain a more comprehensive picture of the systems under consideration. We validated our visualization method and generative approach by qualitatively and quantitatively comparing the results obtained with different rating datasets. Finally, we considered two novel recommendation approaches as possible use cases of multicriteria evaluation methods.

Conclusion In summary, this dissertation deals with the problem of conducting an offline comparison of recommender systems considering both traditional and sequential scenarios. We designed the proposed frameworks for addressing the most critical problems that emerged from literature, namely the non-reproducibility of the results, the comparability of different studies, and the bias of relying on a few metrics only. The software code of our frameworks is freely available on the Web in an attempt to foster further reuse and extension.