

Evaluation Metrics Trade-off for Recommendation Systems

Keywords: recommendation systems, machine learning, big data, evaluation.

Supervisor: Aleksei Triastcyn {firstname.lastname@epfl.ch}

Description

Following the famous Netflix Prize competition, accuracy metrics (such as RMSE) became a *de facto* standard in the community to evaluate and compare performance of recommendation algorithms. However, it has been shown, that accuracy is not the best way to quantify the quality of recommendations, especially in the context of predicting online performance from offline evaluations [1]. Thus, it makes sense to look at other metrics, such as novelty, coverage, or serendipity, and in particular, trade-off between these metrics [2].

In this project, a student will work with large news recommendation datasets, collected online on *swissinfo.ch* and *lepoint.fr*. The goal would be to implement a number of recommender algorithms with tunable hyperparameters, and generate trade-offs between different evaluation metrics by varying these hyperparameters. A student will then analyse these trade-offs by plotting trade-off curves, comparing them for various algorithms, examining the area under the curves, and possibly looking at their impact on CTR and online success rate.

Tasks

- Get familiar with modern recommendation techniques and implement them.
- Get familiar with evaluation metrics and prior work on the topic.
- Analyse trade-off between accuracy and other metrics, and potentially its effect on online performance.

References

- [1] Garcin, F., Faltings, B., Donatsch, O., Alazzawi, A., Bruttin, C., Huber, A., *Offline and online evaluation of news recommender systems at swissinfo.ch*, In Proceedings of the 8th ACM Conference on Recommender systems, pp. 169-176, 2014.
- [2] Maksai, A., Garcin, F., Faltings, B., *Predicting Online Performance of News Recommender Systems Through Richer Evaluation Metrics*, In Proceedings of the 9th ACM Conference on Recommender Systems, pp. 179-186, 2015.