

Abstract of the doctoral thesis

„Collaborative filtering recommender systems in online classifieds”

In many practical applications, the number of available products often exceeds what users can feasibly familiarize themselves with, given the time they are willing to spend. Recommendation systems address this problem by selecting a few items to be presented to each user. By facilitating user decisions, these systems have a significant impact on online classifieds. Commonly used are collaborative filtering models that solely utilize information about interactions between users and items. In the dissertation, we present the results of a study on collaborative filtering models in online classifieds, using OLX Group's websites as examples.

In this work, we discuss the specific characteristics of online classifieds that influence the selection of recommendation and evaluation methods. We present offline evaluation results for existing models and show the advantage of the RP3Beta model over other methods. Furthermore, we verify these results by conducting A/B tests involving millions of users of the OLX Group's websites.

We present the infrastructure created to generate recommendations in real-time and show its applicability to many existing recommendation models. We then propose the RP3Beta real-time model that utilizes this infrastructure. We present the results of A/B tests showing a statistically significant advantage of this model over the baseline model.

We introduce a new graph-based recommendation method, P3LTR, which is a generalisation of the RP3Beta model. The proposed method, unlike the RP3Beta model, has parameters that are optimised when training the model and allows including features of interactions, users and items. We demonstrate the advantage of the P3LTR model over the RP3Beta model in terms of accuracy and coverage metrics during offline evaluation.

We also propose a new graph neural network, P3GNN. We show its superiority in terms of accuracy metrics over existing graph neural networks on the considered dataset. The proposed network achieves similar values of accuracy metrics to the RP3Beta model but allows the generation of vector representations of vertices, which is useful for numerous applications. In addition, there are many elements of the P3GNN network that can be further improved.

We also present a dataset published by the author that allows other researchers to reproduce parts of the reported results using the published source code. We point out several advantages of this dataset over other datasets used to evaluate job recommendation models.

Additionally, we describe the details of the process of successful implementation of the discussed solutions on OLX Group's websites, including an evaluation of their impact on business metrics.



Robert Kwieceński