# Cold-Start Collaborative Filtering

Xiaoxue Zhao
University College London
*x.zhao@cs.ucl.ac.uk*

8 March 2016

**Abstract**

Collaborative Filtering (CF) is a technique to generate personalised recommendations for a user from a collection of correlated preferences in the past. In general, the effectiveness of CF greatly depends on the amount of available information about the target user and the target item. The cold-start problem, which describes the difficulty of making recommendations when the users or the items are new, remains a great challenge for CF. Traditionally, this problem is tackled by resorting to an additional interview process to establish the user (item) profile before making any recommendations. During this process the user's information need is not addressed. In this thesis, however, we argue that recommendations would be preferably provided right from the beginning. And the goal of solving the cold-start problem should be maximising the overall recommendation utility during all interactions with the recommender system. In other words, we should not distinguish between the information-gathering and recommendation-making phases, but seamlessly integrate them together. This mechanism naturally addresses the cold-start problem as any user (item) can immediately receive sequential recommendations without providing extra information beforehand.

This thesis solves the cold-start problem in an interactive setting by focusing on four interconnected aspects. First, we consider a continuous sequential recommendation process with CF and relate it to the exploitation-exploration (EE) trade-off. By employing probabilistic matrix factorization, we obtain a structured decision space and are thus able to leverage several EE algorithms, such as Thompson sampling and upper confidence bounds, to select items. Second, we extend the sequential recommendation process to a batch mode where multiple recommendations are made at each interaction stage. We specifically discuss the case of two consecutive interaction stages, and model it with the partially observable Markov decision process (POMDP) to obtain its exact theoretical solution. Through an in-depth analysis of the POMDP value iteration solution, we identify that an exact solution can be abstracted as selecting users (items) that are not only highly relevant to the target according to the initial-stage information, but also highly correlated with other *potential* users (items) for the next stage. Third, we consider the intra-stage recommendation optimisation and focus on the problem of personalised item diversification. We reformulate the latent factor models using the mean-variance analysis from the portfolio theory in economics. The resulting portfolio ranking algorithm naturally captures the user's interest range and the

uncertainty of the user preference by employing the variance of the learned user latent factors, leading to a diversified item list adapted to the individual user. And, finally, we relate the diversification algorithm back to the interactive process by considering inter-stage joint portfolio diversification, where the recommendations are optimised jointly with the user's past preference records.

The pdf version of my thesis is available for download at http://discovery.ucl.ac.uk/1474118/.