Learning for Top-N Recommendations | High-Dimensional and Heterogeneous Information
Y. Chen
Top-N recommendations have been widely adopted to recommend ranked lists of items so as to help users identify the items that best fit their personal tastes. Collaborative filtering (CF) has been widely studied to generate recommendations by utilizing the information of user’s historical interactions with items (a.k.a. ratings). However, due to the high-dimensionality of ratings in practical applications with a large number of users and items, existing CF-based methods are facing severe challenges in terms of scalability. The sparsity of ratings caused by the high-dimensionality further challenges the performance of recommendation. Typically, additional auxiliary information associated with users or items (a.k.a. side information) is exploited to overcome the rating sparsity. Unfortunately, in recent multimedia scenarios, such information is also high-dimensional. Besides ratings and side information, other information that is relevant for recommendation is also collected from different data sources. How to effectively integrate such heterogeneous information while preserving the effectiveness of CF is also a challenge. In this thesis, we research on top-N recommendations by learning from high-dimensional information and heterogeneous information, based on which we divide the thesis into two parts.

In the first part of the thesis, we focus on leveraging high-dimensional information for top-N recommendations. The first part contains three research chapters, where we respectively utilize high-dimensional ratings (Chapter 2) and high-dimensional side information (Chapter 3 and 4). In Chapter 2, we propose a new regularization term for item-based collaborative filtering (ICF) to overcome issues brought by the high-dimensionality of ratings. In Chapter 3, we propose a joint learning method that simultaneously performs dimension reduction on high-dimensional side information and estimates parameters of an ICF model. In Chapter 4, we propose a new network structure on top of variational auto-encoder to denoise and harness high-dimensional side information.

In the second part of the thesis, we focus on integrating heterogeneous information for top-N recommendations. The second part also contains three chapters, where we respectively combine ratings with item features (Chapter 5), user behavior with content features (Chapter 6) and generic heterogeneous features (Chapter 7). In Chapter 5, we study the problem of recommending top-N new items by estimating local and global similarity functions that calculate item similarities based on item features. We form a Bayesian generative model to seamlessly integrate item-based collaborative filtering with user clustering and deep learning. In Chapter 6, we work on the problem of reranking research paper recommendations by designing a hybrid reranking model. The proposed method combines information behind user multiple behaviors (click, browse, download and etc.) and paper content features (author, venue, entity and etc.). In Chapter 7, we take the effectiveness of factorization machines (FMs) to utilize heterogeneous information for top-N recommendations. We address the high-dimensionality of feature interactions by designing a Bayesian variable selection model. We propose Bayesian personalized feature interaction selection (BP-FIS) as a framework to select personalized feature interactions for FMs.