An Optimal Scaling Approach to Collaborative Filtering using Categorical Principal Component Analysis

Angelos Markos[1], Manolis Vozalis[2], and Konstantinos Margaritis[2]

[1] Department of Primary Education, Democritus University of Thrace
amarkos@eled.duth.gr
[2] Department of Applied Informatics, University of Macedonia
{mans,kmarg}@uom.gr

Abstract. Collaborative Filtering (CF) is a popular approach employed by Recommender Systems, a term used to describe intelligent techniques that generate personalized recommendations. Among the most common and successful approaches to CF are latent factor models, which directly profile both users and products, and neighborhood models, which analyze similarities between products or users. Latent factor models can tackle two fundamentals problems of CF, data sparsity and scalability, and have received considerable attention in recent literature. In this work, we present an optimal scaling approach to address both of these problems using Categorical PCA (CatPCA) for the low-rank approximation of the user-item ratings matrix, followed by a neighborhood formation step. CatPCA is based on the alternating least squares (ALS) algorithm utilizing an optimal scaling process where original data are transformed so that their overall variance is maximized. Unlike other matrix factorization methods, CatPCA does not assume multivariate normality and linear relationships between variables and provides a flexible framework for parameterization. We considered different analysis levels for the variables, starting with a multiple nominal and consecutively applying nominal, ordinal and numeric restrictions. Experiments were executed on the MovieLens dataset, aiming to evaluate the aforementioned methods in terms of accuracy. Results indicated that a combined approach (discretization, multiple nominal scaling level, passive missing value imputation) clearly outperforms the other tested options.

Keywords: Collaborative filtering, Recommender systems, Low-rank approximation, Categorical Principal Component Analysis, Optimal scaling