Statistica Sinica **33** (2023), 1-23
doi:https://doi.org/10.5705/ss.202020.0243

ROBUST RECOMMENDATION VIA SOCIAL NETWORK ENHANCED MATRIX COMPLETION

Jingxuan Wang[1], Haipeng Shen[2] and Fei Jiang[1]

[1] *University of California, San Francisco and* [2] *University of Hong Kong*

Abstract: Robust product recommendation enables internet platforms to boost their business. However, in practice, the user–product rating matrix often has many missing entries. Social network information generates new insights about user behaviors. To fully use such information, we develop a novel approach, called MCNet, that combines the random dot product graph model and low-rank matrix completion to recover missing entries in a user–product rating matrix. Our algorithm improves the accuracy and the efficiency of recovering the incomplete matrices. We study the asymptotic properties of the estimator. Furthermore, we perform extensive simulations and show that MCNet outperforms existing approaches, especially when the data have small signals. Moreover, MCNet yields robust estimation under misspecified models. We apply MCNet and its competitors to predict the missing entries in the user–product rating matrices of the Yelp and Douban movie platforms. The results show that, in general, MCNet gives the smallest testing errors among the comparative methods.

Key words and phrases: Low-rank estimation, matrix completion, missing data, random dot product graph, social network.

1. Introduction

Uncovering true user ratings on products is critical for internet platforms such as Yelp, Facebook, and Amazon, because they help to promote their business. These platforms use estimated ratings to recommend products to users with the highest willingness to pay, thus maximizing their revenue. Such data sets are often arranged in matrix form, where the rows and columns correspond to users and products, respectively. However, typically, many entries are missing, because not every product has been exposed to all users. The ratings of these missing entries are crucial to the recommendation strategies of the products on the platforms.

Many matrix completion algorithms have been developed to recover missing entries in a user–product rating matrix, often adopting low-rank estimation frameworks. Srebro, Rennie and Jaakkola (2005) developed a matrix factoriza-

Corresponding author: Fei Jiang, Department of Epidemiology & Biostatistics, University of California, San Francisco, San Francisco, CA 94158, USA. E-mail: fei.jiang@ucsf.edu.