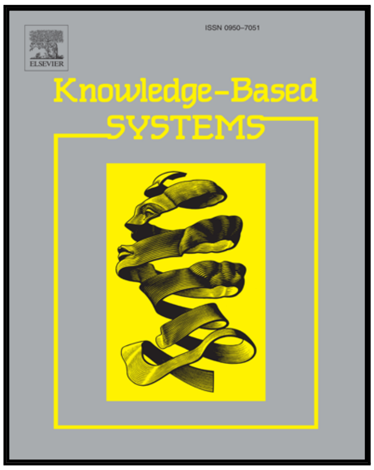
Accepted Manuscript



Performance of Latent Factor Models with Extended Linear Biases

Jia Chen , Xin Luo , Ye Yuan , Mingsheng Shang , Zhong Ming , Zhang Xiong

PII: S0950-7051(17)30067-9
DOI: 10.1016/j.knosys.2017.02.010
Reference: KNOSYS 3823

To appear in: *Knowledge-Based Systems*

Received date: 19 February 2016
Revised date: 6 February 2017
Accepted date: 7 February 2017

Please cite this article as: Jia Chen , Xin Luo , Ye Yuan , Mingsheng Shang , Zhong Ming , Zhang Xiong , Performance of Latent Factor Models with Extended Linear Biases, *Knowledge-Based Systems* (2017), doi: 10.1016/j.knosys.2017.02.010

Performance of Latent Factor Models with Extended Linear Biases

Jia Chen[1, #], Xin Luo[2, 3, #, *], Ye Yuan[2], Mingsheng Shang[2], Zhong Ming[3], Zhang Xiong[1]

1. *School of Computer Science, Beihang University, Beijing 100191, China*

2. *Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China*

3. *College of Computer Science and Engineering, Shenzhen University, Shenzhen 518060, China*

Abstract: High-dimensional and sparse (HiDS) matrices are frequently encountered in various industrial applications, due to the exploding number of involved entities and great needs to describe the relationships among them. Latent factor (LF) models are highly effective and efficient in extracting useful knowledge from these HiDS matrices. They well represent the known data of a HiDS matrix with high computational and storage efficiency. When building an LF model, the incorporation of linear biases has proven to be effect in further improving its performance on HiDS matrices in many applications. However, prior works all propose to assign a single bias to each entity, i.e., a single bias for each user/movie from a user-movie HiDS matrix. In this work we argue that to extend the linear biases, i.e., to assign multiple biases to each involved entity, can further improve an LF model's performance in some applications. To verify this hypothesis, we first extended the linear biases of an LF model, and then deduced the corresponding training rule of involved LFs. Subsequently, we conducted experiments on ten HiDS matrices generated by different industrial applications, evaluating the resulting LF models' prediction accuracy for the missing data of involved HiDS matrices. The experimental results indicate that on most testing cases an LF model needs to extend its linear biases to achieve the highest prediction accuracy. Hence, the number of linear biases should be chosen with care to make an LF model achieve the best performance in practice.

Key Words: *Big Data; Recommender System; Latent Factor Model; Linear Bias; Industrial Application*

1. Introduction

With the world-wild-web widely expanding with a mass of information released, shared and accessed at every moment, people are living in the era of information explosion. They are no longer lack of information but inundated by it. With such information overload, numerous entities and their corresponding high-dimensional relationships are frequently processed in many industrial applications, e.g., users, items and user-item preferences in recommender systems [1-5], users, services and user-service QoS history in Web-service QoS analysis systems [6-8], users and user-user trust networks in social network-based services [9-13], and proteins and protein-protein interactome mappings in bioinformatics [14-15]. Due to the exploding numbers of involved entities, e.g., hundreds of thousands of users and items in recommender systems, it becomes impossible to observe the full relationship among them. Consequently, high-dimensional and sparse (HiDS) matrices are frequently encountered to describe such relationships in practice [1-3, 6-12, 14-15].

In spite of their incompleteness, these HiDS matrices contain rich information regarding various desired patterns, e.g., user preferences in recommender systems [1-3] and protein interactome features

[#]J. Chen and X. Luo contributed equally to this work.

[*]X. Luo is the corresponding author.

E-mail Addresses: chenjia@buaa.edu.cn (J. Chen), luoxin21@cigit.ac.cn (X. Luo), yuanye@cigit.ac.cn (Y. Yuan), msshang@cigit.ac.cn (M. Shang), mingz@szu.edu.cn (Z. Ming), xiongz@buaa.edu.cn (Z. Xiong).