



A goodness-of-fit test on the number of biclusters in a relational data matrix

Chihiro Watanabe¹ · Taiji Suzuki^{1,2}

Received: 1 June 2022 / Accepted: 28 February 2023 / Published online: 17 April 2023
© The Institute of Statistical Mathematics, Tokyo 2023

Abstract

Biclustering is a method for detecting homogeneous submatrices in a given matrix. Although there are many studies that estimate the underlying bicluster structure of a matrix, few have enabled us to determine the appropriate number of biclusters. Recently, a statistical test on the number of biclusters has been proposed for a regular-grid bicluster structure. However, when the latent bicluster structure does not satisfy such regular-grid assumption, the previous test requires a larger number of biclusters than necessary for the null hypothesis to be accepted, which is not desirable in terms of interpreting the accepted structure. In this study, we propose a new statistical test on the number of biclusters that does not require the regular-grid assumption and derive the asymptotic behavior of the proposed test statistic in both null and alternative cases. We illustrate the effectiveness of the proposed method by applying it to both synthetic and practical data matrices.

Keywords Biclustering · Submatrix detection · Goodness-of-fit test · Random matrix theory

✉ Chihiro Watanabe
chihiro.watanabe.xz@hco.ntt.co.jp

¹ Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

² Center for Advanced Intelligence Project (AIP), RIKEN, Tokyo, Japan