A Fused Latent and Graphical Model for Multivariate Binary Data

Yunxiao Chen, Xiaoou Li, Jingchen Liu, Zhiliang Ying

We consider modeling, inference, and computation for analyzing multivariate binary data. We propose a new model that consists of a low dimensional latent variable component and a sparse graphical component. Our study is motivated by analysis of item response data in cognitive assessment and has applications to many disciplines where item response data are collected. Standard approaches to item response data in cognitive assessment adopt the multidimensional item response theory (IRT) models. However, human cognition is typically a complicated process and thus may not be adequately described by just a few factors. Consequently, a low-dimensional latent factor model, such as the multidimensional IRT models, is often insufficient to capture the structure of the data. The proposed model adds a sparse graphical component that captures the remaining ad hoc dependence. It reduces to a multidimensional IRT model when the graphical component becomes degenerate. Model selection and parameter estimation are carried out simultaneously through construction of a pseudo-likelihood function and properly chosen penalty terms. The convexity of the pseudo-likelihood function allows us to develop an efficient algorithm, while the penalty terms generate a low-dimensional latent component and a sparse graphical structure. Desirable theoretical properties are established under suitable regularity conditions. The method is applied to the revised Eysenck's personality questionnaire, revealing its usefulness in item analysis. Simulation results are reported that show the new method works well in practical situations.