Fast Variational Inference for Joint Mixed Sparse Graphical Models

Qingyang Liu and Yuping Zhang

Abstract—**Mixed graphical models are widely implemented to capture interactions among different types of variables. To simultaneously learn the topology of multiple mixed graphical models and encourage common structure, people have developed a variational maximum likelihood inference approach, which takes advantage of the log-determinant relaxation. In this article, we further improve the computational efficiency of this method by exploiting the block diagonal structure of the solution. We present a simple necessary and sufficient condition to identify the connected components in the solution, so as to determine the block diagonal structure. Then, utilizing the idea of "divide-and-conquer", we are able to adapt the joint structural inference problem for multiple related large sparse networks. We illustrate the merits of the proposed algorithm via experimental comparisons in computational speed.**

Index Terms—**Big data, data integration, group lasso, joint modeling, network.**

I. Introduction

In recent biomedical research, reconstruction of interaction networks among a group of features is critical for characterizing their functions and mechanisms, so as to unveil etiology of complex diseases and develop targeted therapies. Undirected graphical model, also known as Markov Random Field, is a popular tool to describe the conditional dependency structure for high-dimensional data. Representative examples include Gaussian graphical models for symmetric and thin-tailed continuous data [1], [2], and Ising models for binary data [3]. To broaden the application of graphical models to data composed of heterogeneous types of features, a line of work gradually developed a subclass of mixed Markov Random Fields [4]–[8], which is specified by conditional distributions belonging to potentially different exponential families. Such mixed graphical models are helpful in modeling complex gene regulatory networks from multi-platform data, for instance, mixtures of gene expressions, mutations, copy number variations, and epigenetic states, including binary, categorical, count, and continuous features.

Depending on the formulation of mixed graphical model, it is more challenging to learn the conditional dependency structure of complex mixtures than Gaussian graphical models. Most current literature, [5]–[7] employed neighborhood selection in their learning algorithms, which may result in asymmetric edge selection results. Other than the common asymmetry problem in nodewise regression, the computation can be inefficient sometimes when some special composite penalties are applied. For instance, [7] used a Newton-type algorithm within each ADMM update, resulting in high algorithmic complexity. To address this problem, [9] proposed a variational inference approach using the log-determinant relaxation. Originally proposed in [10], [11], the log-determinant relaxation replaces the log-partition function by a Gaussian entropy bound, which is a method to approximately marginalize the discrete graphical model. Utilizing the same upper bound to construct an approximate likelihood maximization problem for mixed graphical models, [9] demonstrated that the practical performance of this variational method is satisfactory regarding edge selection.

It is extensively observed that different biological systems share alike regulation mechanisms in many studies. Thus, integrating data from different biological conditions and exploiting the prospective similarity among them is expected to lead to more precise and efficient structural inference. Along this line of research, [12] proposed joint graphical lasso to estimate multiple similar Gaussian graphical models. Reference [13] proposed a data-integration framework to jointly learn multiple mixed graphical models for categorical and Gaussian data, and applied it to copy number variations and mutations from two cancer types. Extending the data integration to more complex models, [14] adopted the aforementioned variational approach to formulate a joint inference for exponential Markov Random Fields and applied the algorithm to data composed of Gaussian, categorical, Poisson, truncated Poisson, or exponential variables.

Among a considerable number of regulatory features, we are

Manuscript received May 15, 2020; revised October 26, 2020 and November 28, 2020; accepted November 28, 2020. Date of publication