A CONSTRAINED $\ell 1$ MINIMIZATION APPROACH FOR ESTIMATING MULTIPLE SPARSE GAUSSIAN OR NON-PARANORMAL GRAPHICAL MODELS

Beilun Wang, Ritambhara Singh & Yanjun Qi
Department of Computer Science University of Virginia, Charlottesville, VA 22901, USA
{bw4mw,rs3zz,yanjun}@virginia.edu

ABSTRACT

The flood of multi-context measurement data from many scientific domains has created an urgent need to reconstruct context-specific variable networks, that could significantly simplify network-driven studies. Computationally, this problem can be formulated as jointly estimating multiple different, but related, sparse Undirected Graphical Models (UGM) from samples aggregated across several contexts. Previous joint-UGM studies could not address this challenge since they mostly focus on Gaussian Graphical Models (GGM) and have used likelihood-based formulation to infer multiple graphs toward a common pattern. Differently, we propose a novel approach, SIMULE (learning Shared and Individual parts of MULtiple graphs Explicitly) to solve multi-task UGM using a $\ell 1$ constrained optimization. SIMULE is cast as independent subproblems of linear programming that can be solved efficiently. It automatically infers specific dependencies that are unique to each context as well as shared substructures preserved among all the contexts. SIMULE can handle both multivariate Gaussian and multivariate Nonparanormal data that greatly relax the normality assumption. Theoretically we prove that SIMULE achieves a consistent result at rate $O(\log(Kp)/n_{tot})$ (not been proved before). On four synthetic datasets, SIMULE shows significant improvements over state-of-the-art multi-sGGM and single-UGM baselines. [1]

1 INTRODUCTION

In this paper, we consider the problem of jointly estimating K undirected graphical models (UGM) from K multivariate sample blocks. Each data block contains a different set of data observations, on the same set of variables. This is motivated by the fact that the past decade has seen a revolution in collecting large-scale heterogeneous data from many scientific fields like genetics or brain science. Given such data, understanding and quantifying variability of variable graphs across multiple contexts is a fundamental analysis task. This is because, for example, quantifying and analyzing interaction edges (representing statistical dependencies) between molecules, that are activated only under a specific context, can predict or help to understand the existence or severity of such a context. In order to address the above issue, we propose a novel approach that uses a $\ell 1$ constrained minimization formulation for joint structure learning of multiple sparse GGMs or sNGMs(sparse nonparanormal Graphical models). We name the method SIMULE (learning Shared and Individual parts of MULtiple graphs Explicitly).Using a $\ell 1$ constrained optimization strategy (Section 2), SIMULE extends CLIME(Cai et al., 2011) to a multi-tasking setting. The learning step is solved efficiently through a formulation of multiple independent sub-problems of linear programming for which we also provide a parallel version of the learning algorithm. Compared with previous multi-task sGGM models, SIMULE can accurately quantify task-specific network variations that are unique for each task. This also leads to a better generalization and benefits all the involved tasks.

We provide evaluations through both experimental results and theoretical analysis. Theoretically we prove that SIMULE and its nonparanormal extension NSIMULE achieve a consistent estimation of the target (true) dependency graphs with a high probability at rate $O(\log(Kp)/n_{tot})$. Here n_{tot} represents the total number of samples from all tasks and K describes the number of tasks. This proof also theoretically validates the benefit of learning multiple sGGMs jointly, since the $O(\log(Kp)/n_{tot})$ convergence rate is better than learning multiple single-sGGMs separately whose

[1] An extended version of the work is currently under review at the ECML-PKDD 2016 journal-track.

rate is $O(\log p/n_i)$ respectively. Here n_i represents the number of samples of i-th task. Experimentally, we show strong improved performance of SIMULE and NSIMULE over multiple baseline methods on four synthetic datasets. The proposed methods obtain better AUC and partial AUC scores on all simulated cases.

2 METHOD

SIMULE: Infer Shared and Individual Parts of Mulltiple sGGM Explicitly: Treating sparse GGM estimation from each data block as a single task, our main task is to learn multiple sGGMs over K data blocks or K tasks jointly, which can lead to better generalization and benefit all of the involved tasks (theoretically proved in Section 2.2). First of all, We model each GGM network as:

$$\Omega^{(i)} = \Omega_I^{(i)} + \Omega_S, \tag{1}$$

where Ω_S is the shared pattern among all graphs and $\Omega_I^{(i)}$ represents the individual part specific for the i-th graph.

Secondly, we choose the CLIME estimator to model each sGGM task. Following the idea of joint estimators of multiple graphical lasso (described in Section 2.1), we sum up multiple CLIME estimators from each task for multi-tasking. This gives us the following novel formulation of SIMULE:

$$\Omega_I^{(i)}, \Omega_S = \underset{\Omega_I^{(i)}, \Omega_S}{\operatorname{argmin}} \sum_i |\Omega_I^{(i)}|_1 + K|\Omega_S|_1 \tag{2}$$

$$\text{Subject to: } |\Sigma^{(i)}(\Omega_I^{(i)} + \Omega_S) - I|_\infty \leq \lambda_n,\ i = 1, \ldots, K$$

In Section 2.2, we theoretically prove that the estimated $\Omega^{(i)}$ from Eq. (1) will be positive definite and converge to the true precision matrices with a high probability. We also can solve Eq. (2) column by column without influencing the resulting solution.

Finally, the Eq. (2) is equivalent to a linear programming problem. To solve it, we follow the primal dual interior method (Boyd & Vandenberghe, 2004) that has also been used in the Dantzig selector for regression (Candes & Tao, 2007).

Method Variation: nonparanormal SIMULE (NSIMULE) Though sGGM is powerful, its normality assumption is commonly violated in real applications such as in the biomedical data we try to analyze in Section 3. The nonparanormal and transelliptical Graphical Models recently proposed by (Liu et al., 2012) have extended sGGM to new distribution families. Therefore we extend SIMULE to a novel variation, nonparanormal SIMULE (NSIMULE) that learns to fit multiple nonparanormal graphical models (NGM) (Liu et al., 2009) jointly. The nonparanormal graphical model (NGM) (Liu et al., 2009) assumes that data samples follow a multivariate nonparanormal distribution, which is a strict superset of Gaussian distribution.

2.1 RELATED WORK

The proposed SIMULE extends CLIME (Cai et al., 2011) to a multi-tasking setting. This is different from most previous methods that estimate multiple sGGMs jointly by following the penalized log-likelihood formulation. We choose three most relevant studies that also provide runnable software as our baselines in the experiments: (1) Joint graphical lasso (JGL) (Danaher et al., 2013); (2) Node-perturbed JGL (Mohan et al., 2013); and (3) Simone (Chiquet et al., 2011). Furthermore, (Han et al., 2013) proposed to estimate a population graph from multi-block data using a so-called median-graph idea. It is conceptually similar to the shared part modeled by SIMULE. However, they do not explicitly model individual parts that are specific to each task. Another recent study CSSL-GGM (Hara & Washio, 2013) also tried to model both the shared and individual substructures in multi-sGGMs. Different from ours, their formulation follows the penalized likelihood framework and used ℓ1,p norm to regularize the task-specific parts, while SIMULE uses ℓ1 norm. The ℓ1,p norm they used actually pushes the individual parts of multiple graphs to be similar which is contradictory to the original purpose of these parameters. More recently (Monti et al., 2015) proposed a method to learn population and subject-specific brain connectivity networks via a so-called Mixed Neighborhood Selection (MSN) method. Since MSN is specially designed for brain imaging data, it assumes each individual graph is generated through random effects modeling by latent variables. Our model is more general while MSN may fit brain imaging data better.

2.2 THEORETICAL ANALYSIS

We theoretically prove that we can achieve a good estimation of target dependency graphs with the convergence rate $O(\log(Kp)/n_{tot})$ (details are not shown due to space limitation). We also

prove that the nonparametric assumption would not change this convergence rate. Furthermore, We provide theoretical proofs for the benefits of multi-tasking sGGM. The convergence rate for each task is reduced from $O(\log p/n_i)$ to $O(\log(Kp)/n_{tot})$. Both of these theoretical results haven't been investigated by the previous studies.

3 EXPERIMENT

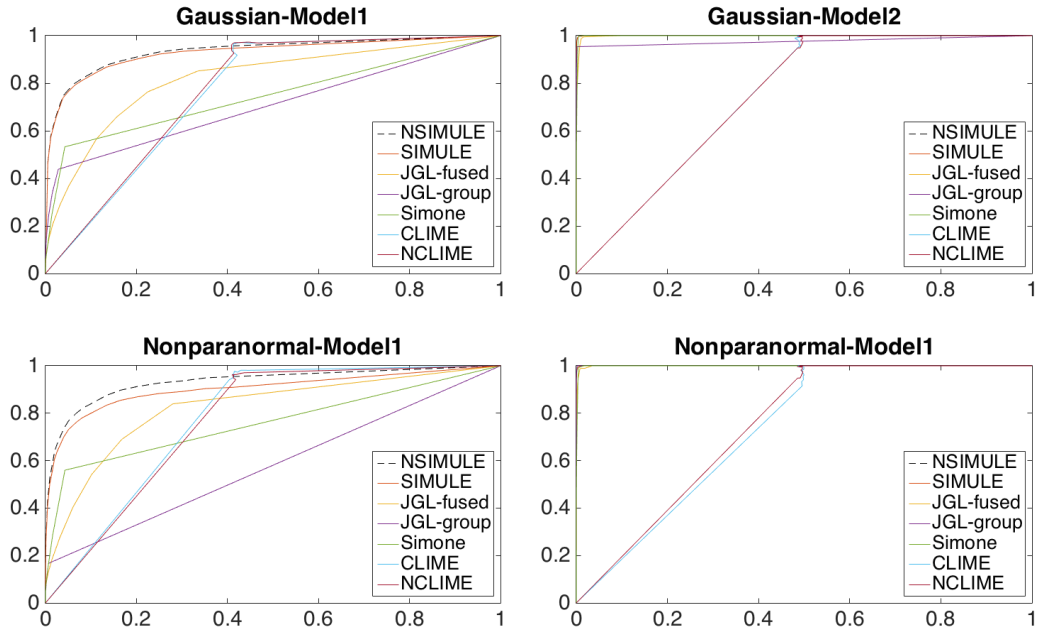

Figure 1: The FPR-TPR curve graph for different methods on four simulated datasets. The upper two are for Gaussian simulated datasets. The lower two are for nonparanormal datasets. The upper left and lower left graphs are generated using graph Model 1. The upper right and lower right are generated using Model 2. Each curve is generated through varying the tuning parameter(s). We can see that curves from SIMULE and NSIMULE are above all baseline methods (more apparent on two datasets generated by Model 1).

In this section, we use four simulated datasets to evaluate the proposed methods. Using two network models[2], we first generate two synthetic multivariate Gaussian datasets and two nonparanormal datasets, in which each model includes K tasks of data samples (We choose $K = 2$ for simplicity). We compare our model with the baseline methods mentioned in Section 2.1. The edge-level false positive rate (FPR) and true positive rate (TPR) are used to measure the difference between the true graphs and the predicted graphs. The FPR vs. TPR curves generated by different methods are provided in Figure 1. The first row of Figure 1 presents FPR vs. TPR plots from SIMULE, NSIMULE and the baseline methods on two simulated Gaussian datasets from Model 1 and Model 2. The subfigure "Gaussian-Model1" clearly shows that our methods obtain better under-plot areas than three multi-sGGM baselines and two single-sGGM baselines. We can also conclude all multi-task estimators perform better than single-CLIME and single-NCLIME estimators. On the subfigure "Gaussian-Model2", all multi-task estimators perform better than single-CLIME and single-NCLIME estimators. The differences among various multi-sGGM estimators are not as apparent as in the subfigure "Gaussian-Model1". Our methods also outperform baselines on two real-world datasets (results are not shown due to space limitation).

[2]Model 1 generates two random sparse graphs. Model 2 provides two structured graphs, i.e., grid and ring. We choose $p = 100$, i.e., the dimension of random variables is 100. For each dataset, 500 data samples are generated randomly. Using either model 1 or model 2, we generate K blocks of data samples with i-th block following $N(0, (\Omega^{(i)})^{-1})$. Then we apply SIMULE, NSIMULE and baseline models to these datasets to obtain the estimated dependency networks.

REFERENCES

Stephen P Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

Tony Cai, Weidong Liu, and Xi Luo. A constrained 1 minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607, 2011.

Emmanuel Candes and Terence Tao. The dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics*, pp. 2313–2351, 2007.

Julien Chiquet, Yves Grandvalet, and Christophe Ambroise. Inferring multiple graphical structures. *Statistics and Computing*, 21(4):537–553, 2011.

Patrick Danaher, Pei Wang, and Daniela M Witten. The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2013.

Fang Han, Han Liu, and Brian Caffo. Sparse median graphs estimation in a high dimensional semiparametric model. *arXiv preprint arXiv:1310.3223*, 2013.

Satoshi Hara and Takashi Washio. Learning a common substructure of multiple graphical gaussian models. *Neural Networks*, 38:23–38, 2013.

Han Liu, John Lafferty, and Larry Wasserman. The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *The Journal of Machine Learning Research*, 10:2295–2328, 2009.

Han Liu, Fang Han, and Cun-hui Zhang. Transelliptical graphical models. In *Advances in Neural Information Processing Systems*, pp. 809–817, 2012.

Karthik Mohan, Palma London, Maryam Fazel, Su-In Lee, and Daniela Witten. Node-based learning of multiple gaussian graphical models. *arXiv preprint arXiv:1303.5145*, 2013.

Ricardo Pio Monti, Christoforos Anagnostopoulos, and Giovanni Montana. Learning population and subject-specific brain connectivity networks via mixed neighborhood selection. *arXiv preprint arXiv:1512.01947*, 2015.