

# 基于多尺度空间相关的气象要素预测

刘丽丹

(南京航空航天大学 计算机科学与技术学院,江苏 南京 210018)

**摘要:**气象要素预测与人们的日常生活密切相关,为了探索气象要素的空间相关性质,进而实现不同时间尺度的气象要素预测,采用了图信号方法,利用能够简洁地刻画变量间依赖关系的概率图模型,由各个气象观测站点数据构成图模型的节点,建立条件高斯图模型,学习图中边的分布规律,以确定站点间的空间相关关系。根据天气系统的特征尺度,对条件高斯图模型进行改进,引入相似度矩阵,通过交叉验证的方法确定各个预测时间尺度所对应的相似度矩阵参数,建立多尺度空间相关模型,利用改进的二阶有效集方法进行模型估计。将此模型应用于全球3 431个国际交换站,分别进行了6小时、24小时、72小时和7天降水量与气温的气象要素预测。实验结果表明,该模型在计算效率和预测准确率方面都有所提升。

**关键词:**概率图模型;条件高斯图模型;相似度矩阵;空间相关;要素预测

**中图分类号:**TP399

**文献标识码:**A

**文章编号:**1673-629X(2019)12-0148-05

**doi:**10.3969/j.issn.1673-629X.2019.12.026

## Meteorological Elements Forecasting Based on Multi-scale Spatial Correlation

LIU Li-dan

(School of Computer Science and Technology, NUAU University, Nanjing 210018, China)

**Abstract:** Meteorological element prediction is closely related to people's daily life. In order to explore the spatial correlation properties of meteorological elements and then realize the prediction of meteorological elements at different time scales, the graph signal method is adopted. The probability graph model which can concisely depict the dependence relationship between variables is used. The node of the graph model is composed of the data of each meteorological observation station, and the conditional Gauss graph model is established. To determine the spatial correlation between sites, the distribution of edges in the exercise is studied. According to the characteristic scale of weather system, the conditional Gaussian graph model is improved, and the similarity matrix is introduced. The parameters of the similarity matrix corresponding to each prediction time scale are determined by cross-validation method. A multi-scale spatial correlation model is established, and the model is estimated by using the improved second-order effective set method. The model is applied to 3 431 international exchange stations in the world, and the meteorological elements of precipitation and temperature are predicted for 6 hours, 24 hours, 72 hours and 7 days respectively. The experimental shows that the model has improved both in calculation efficiency and prediction accuracy.

**Key words:** probability graph model; conditional Gauss graph model; similarity matrix; spatial correlation; factor prediction

## 0 引言

空间相关,是指一些变量在同一个分布区内的观测数据之间潜在的相互依赖性。如果是同一个变量在不同空间位置的变化存在相互依赖关系,称为空间自相关,不同的变量之间的空间相关则为空间互相关。气象要素的空间自相关反映的是一个站点观测到的要素值与附近相邻站点的同一种要素值的相关程度,是

一种检测与量化各站点间该要素的空间依赖性的空间统计方法。由于大气的连续性,气象要素的空间自相关性是比较明显的,比如相邻地区的海平面气压值,一般都是平稳变化的。但由于受观测站地形的影响,某些要素值的空间自相关性也会表现的不同,比如山顶和山脚两个测站的气温,有海拔高度的因素,也有日照的影响,单从水平空间距离考虑,空间相关性可能就不

收稿日期:2019-01-08

修回日期:2019-05-08

网络出版时间:2019-06-27

基金项目:国家自然科学基金青年科学基金(41405135)

作者简介:刘丽丹(1981-),女,工程师,CCF会员(A5938G),研究方向为智能计算与机器学习。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190627.1105.056.html>

明显。

概率图模型是用图形来表达变量之间相关关系的概率模型,可以简洁地刻画出复杂分布的结构,具有表示、推理和学习的能力<sup>[1]</sup>,现已被广泛应用于机器学习、计算机视觉、自然语言处理、语音识别、专家系统、用户推荐、社交网络挖掘、生物信息学等研究领域的最新成果中<sup>[2-3]</sup>。概率图模型可以根据边的性质分为两类:一类是使用有向无环图表示变量之间的依赖关系,称为有向图模型或贝叶斯网;另一类是使用无向图表示变量之间的相关关系,称为无向图模型或马尔可夫网<sup>[4]</sup>。目前,无向图模型的主要应用包括图像重构、图像分割、图像恢复、3D 视觉、目标识别和目标匹配等<sup>[5]</sup>。高斯图模型是概率图模型的一种,是无向图模型,它假设变量服从高斯分布。高斯图模型虽带有强假设,但在数学上易于处理,而且根据中心极限定理,可以较好地近似实际连续分布<sup>[6]</sup>,可用于异常目标检测<sup>[7]</sup>和图像分割<sup>[8-9]</sup>等领域。而条件高斯图模型,不仅能模拟随机变量之间的条件相关关系,也能模拟随机变量与协变量之间的条件相关关系。

对于条件高斯图模型来说,边的估计,即是估计其协方差矩阵,亦是估计其精度矩阵。对于精度矩阵的估计,比较常用的方法有 Lasso、牛顿法、坐标下降和梯度下降等等。1996 年 Lauritzen 等用 greedy forward-backward search 方法来估计精度矩阵<sup>[10]</sup>,但被认为在计算上并不可行。Friedman 等提出了 Graphical Lasso 算法,以 L1 正则项控制精度矩阵的稀疏度,用最大似然来估计精度矩阵<sup>[11]</sup>。后来又不断涌现出多种方法,比如:梯度下降算法<sup>[12]</sup>、平滑优化法<sup>[13]</sup>、块坐标下降法<sup>[14]</sup>、交替线性优化法<sup>[15]</sup>、二次逼近法<sup>[16]</sup>等等。Matt Wytock 提出了一种二阶有效集的方法求解稀疏条件高斯图模型,通过迭代得到不带 L1 惩罚项的目标函数的二阶近似,然后求解一个 L1 正则化二次规划来找到牛顿步长的正则化模拟。这种方法在大规模问题中更加实用,速度也更快<sup>[17]</sup>。

由于天气系统的特征尺度不同,在不同的时间尺度上,所需的验证空间也是有区别的。如果要预测 6 小时后某一站点的气象要素,和它相关的或者说是能够影响预测结果的站点,与该站点的空间距离应被限定在一个较小的范围,而如果预测时效为 7 天,则空间相关的距离就应该相应扩大。当然天气或要素的模拟推断并非如此简单,还有天气系统间的相互作用必须要考虑,在各种尝试之后,根据预测时效来调整空间距离影响的权重,在模型初始化阶段引入相似度矩阵,建立多尺度空间相关模型(MS-CGGM)。一方面能够提高模型计算效率,另一方面也可以达到去噪效果,提高预测精度。

1 多尺度空间相关图模型

设输出变量为  $Y \in R^{n \times p}$ ,输入变量为  $X \in R^{n \times q}$ ,  $q = m \times p$ ,  $n$  为样本量,  $p$  为站点数,  $m$  为协变量维数,即参与预测的气象要素个数。利用条件高斯图模型,有:

$$Y = f(X) = -C^{-1}\Theta^{-1}X \tag{1}$$

条件分布:  $P(y | x; C, \Theta) = \frac{1}{Z(x)}e^{-y^TCy - 2x^T\Theta y}$ , 其中  $Z(x) = \pi^{p/2} |C|^{-1}e^{x^T\Theta C^{-1}\Theta x}$  为划分函数。  
样本对  $\{y_i, x_i\}_{i \in n}$  的对数似然为:

$$f(C, \Theta) = -\sum_{i=1}^n \log P(y_i | x_i; C, \Theta) = -\log |C| + \text{tr}(S_{yy}C + 2S_{yx}\Theta + (C)^{-1}(\Theta)^TS_{xx}\Theta) \tag{2}$$

其中,  $S_{yy} = \frac{1}{n}Y^TY$ 、 $S_{yx} = \frac{1}{n}Y^TX$ 、 $S_{xx} = \frac{1}{n}X^TX$  为样本协方差矩阵。

应用 Matt Wytock 的估计方法<sup>[17]</sup>,最小化负对数似然并加入  $\ell_1$  惩罚项,即:

$$\min_{C, \Theta} f(C, \Theta) + \lambda (\|C\|_1 + \|\Theta\|_1) \tag{3}$$

根据一阶非光滑优化方法,  $f(C, \Theta)$  的梯度定义为:

$$\begin{cases} \nabla_C f(C, \Theta) = S_{yy} - C^{-1} - C^{-1}\Theta^TS_{xx}\Theta C^{-1} \\ \nabla_\Theta f(C, \Theta) = 2S_{yx} + 2S_{xx}\Theta C^{-1} \end{cases} \tag{4}$$

在每次迭代中使用有效集方法,如果

$$\begin{cases} |(\nabla_C f(C, \Theta))_{i,j}| > \lambda \text{ 或 } C_{ij} \neq 0 \\ |(\nabla_\Theta f(C, \Theta))_{i,j}| > \lambda \text{ 或 } \Theta \neq 0 \end{cases} \tag{5}$$

每次只优化一个  $(\Delta_C)_{i,j}$  或者  $(\Delta_\Theta)_{i,j}$ 。

由式 6 得到牛顿加速方向:

$$D_C, D_\Theta \leftarrow \underset{\Delta_C, \Delta_\Theta, \Delta_i=0}{\text{argmin}} g(C, \Delta_\Theta) + \lambda (\|C + \Delta_C\|_{1,*} + \|\Theta\|_1) \tag{6}$$

二阶有效集方法地流程如下:

- (1) 给定  $X, Y, \lambda$ , 初始化  $C = I, \Theta = 0$ ;
- (2) 用式 5 计算有效集  $S_C, S_\Theta$ 。
- (3) 根据式 6 寻找牛顿加速方向  $D_C, D_\Theta$ ;
- (4) 用回溯线搜索法 (backtracking line search) 计算步长  $\alpha$  (期间要保证  $C$  是正定的);
- (5) 迭代:  $C \leftarrow C + \alpha D_C, \Theta \leftarrow \Theta + \alpha D_\Theta$ ;
- (6) 返回步骤 2。

由此得到  $C$  和  $\Theta, C$  即为空间相关关系矩阵。

根据如上所述,引入一个相似度矩阵  $D$ , 根据天气的特征尺度对预测模型施加约束。

$$D = \begin{bmatrix} d_{11} & \cdots & d_{1j} & \cdots & d_{1p} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{i1} & \cdots & d_{ij} & \cdots & d_{ip} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{p1} & \cdots & d_{pj} & \cdots & d_{pp} \end{bmatrix}$$

其中,  $d_{ij}$  取拉普拉斯核  $d_{ij} = \exp(-\frac{\|s_i - s_j\|}{b})$ ,  $b > 0$ ,  $s_i, s_j$  代表第  $i$  和第  $j$  个站点的地理位置, 带宽  $b$  则由预测时效决定。采用交叉验证的方法根据预测时效来找到最佳带宽。

式 2 中  $S_{yy} = \frac{1}{n} Y^T Y, S_{yx} = \frac{1}{n} Y^T X, S_{xx} = \frac{1}{n} X^T X$  变为:  $S'_{yy} = \frac{1}{n} D \odot Y^T Y, S'_{yx} = \frac{1}{n} D \odot Y^T X, S'_{xx} = \frac{1}{n} D \odot X^T X$ , 其中  $\odot$  表示 Hadamard 积, 为两个矩阵对应元素的乘积。

式 2 转换为:

$$f(C, \Theta) = - \sum_{i=1}^n \log P(y_i | x_i; C, \Theta) = - \log |C| + \text{tr}(S'_{yy} C + 2S'_{yx} \Theta + (C)^{-1} (\Theta)^T S'_{xx} \Theta)$$

(7)

再应用二阶有效集的方法估计  $C$  和  $\Theta$ , 得到空间相关关系矩阵。

2 实验与结果

2.1 实验数据与评估标准

采用全球国际交换站 2003—2012 共 10 年的定时气象观测数据, 来源于全球地面天气资料定时值数据集。由于一些站点数据不全, 经过筛选, 选取了遍布全球各个州的 3 431 个站点, 将数据进行了清理、归一化和中心化, 以前 9 年数据为训练数据集, 2012 年数据作为测试数据集。

2.2 实验结果

为了学习全球范围的观测站点之间的空间相关关系, 尝试了不同的方法。空间范围大至全球后, 简单的条件高斯图模型处理起来有些困难, 迭代次数明显增多, 学习出的空间关系参考价值也不明显。图 1 为第 1—50 号站点的空间相关系数图, 其中各图按预测时效区分, 图 a 是 6 小时, 图 b 是 24 小时, 图 c 是 72 小时, 图 d 是 7 天。图中坐标示数代表站点。

如果直接使用天气系统的特征尺度设定一个空间相关边界, 效果也不理想, 预测误差明显变大。图 2 是人为设定相关边界的空间相关系数图, 左图为 24 小时降水空间相关系数, 右图为 6 小时气温空间相关系数图, 图中坐标示数代表站点。这种强假设带来了相关系数图中过于清晰的边界, 忽略了天气系统之间的相互作用, 并且到底哪种距离范围更合适也很难界定, 所以此种方法并不可取。

应用多尺度空间相关模型, 先用交叉验证法根据预测时效选择代表带宽的参数  $b$ , 选择使得误差最小的带宽, 得到表 1 所示的时间-空间尺度约束。

表 1 不同预测时效对应的最优核函数带宽

预测时效	最优带宽 b/km
6 小时	900
24 小时	1 700
72 小时	2 000
7 天	2 200

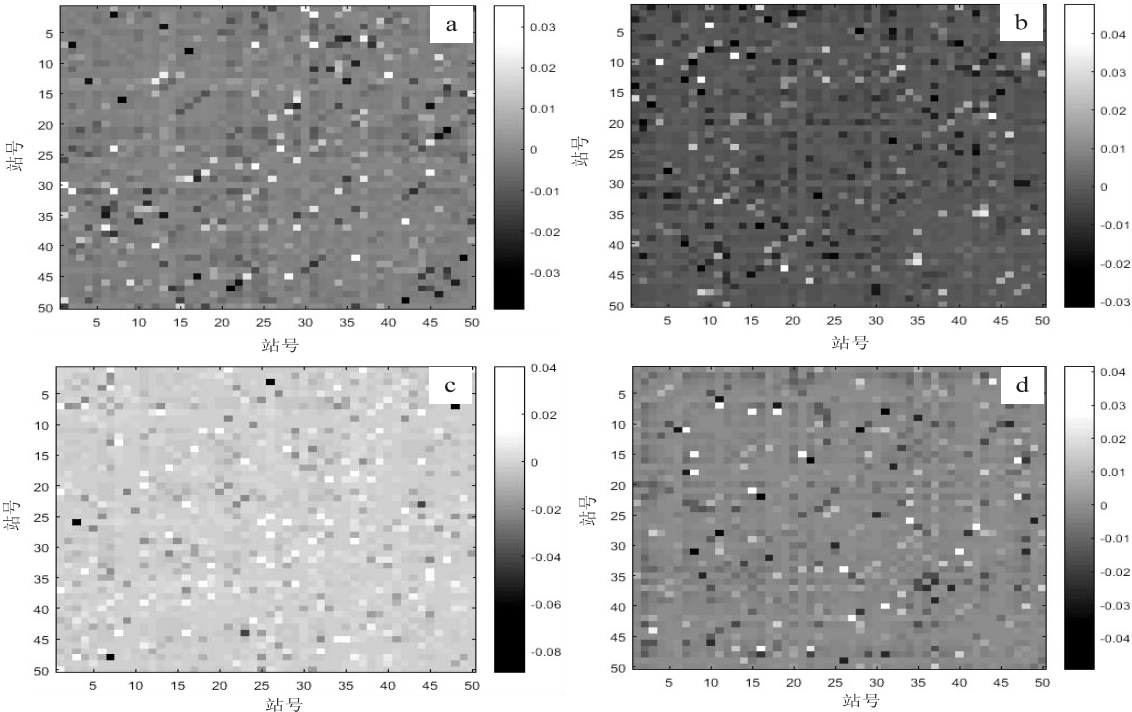


图 1 CGGM 得到的 1—50 号站点不同时效空间相关系数图

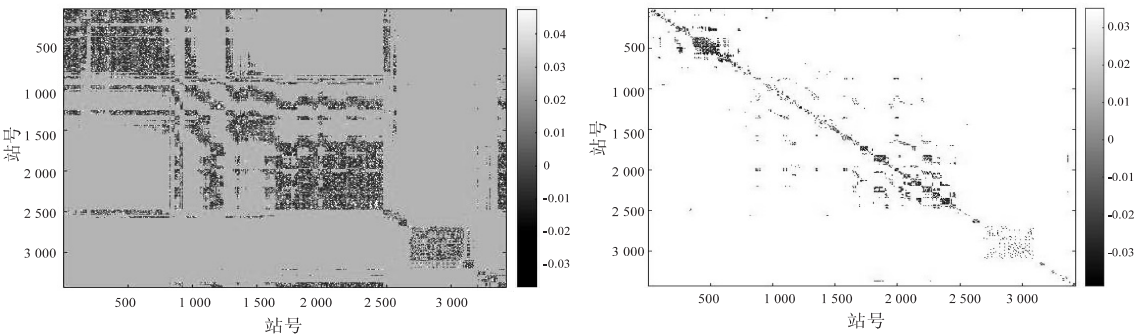


图 2 带相关边界约束的空间相关系数图

利用多尺度条件高斯图模型对全球 3 431 个站点进行训练,得到了不同预测时效下的空间相关关系。图 3 是第 1—50 号站点的空间相关系数图,其中各图仍然按预测时效区分,图 a 是 6 小时,图 b 是 24 小时,图 c 是 72 小时,图 d 是 7 天。图中坐标示数代表站点。

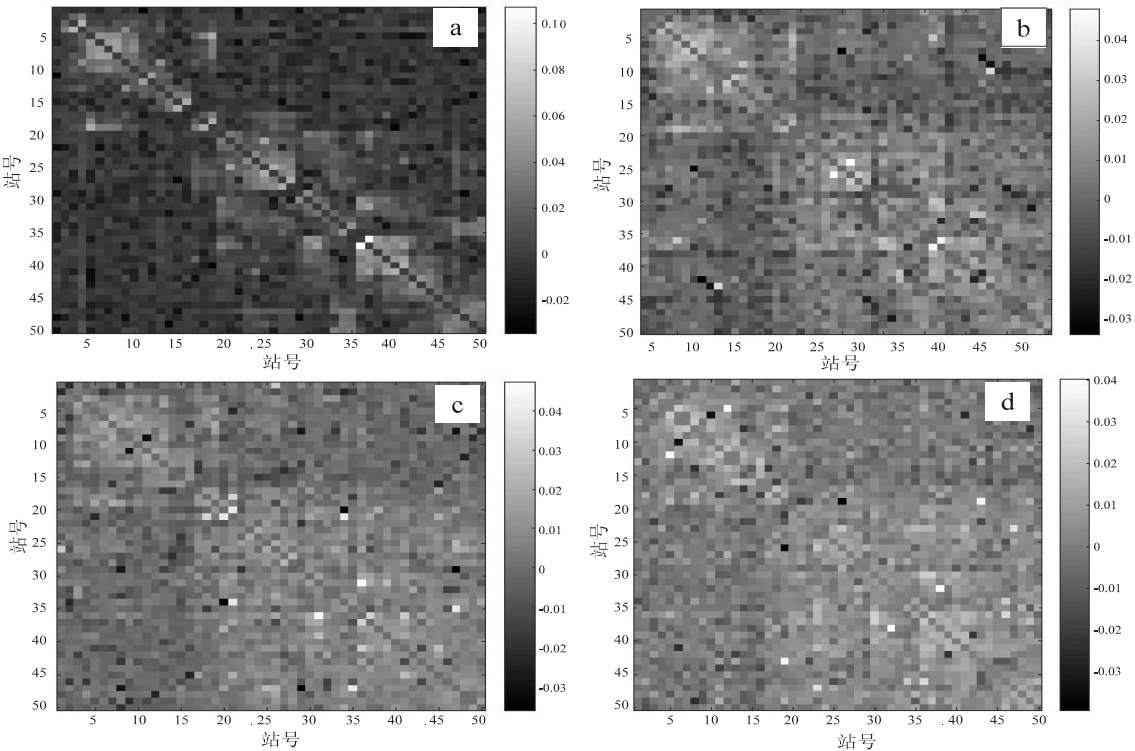


图 3 MS-CGGM 得到的 1-50 号站点不同时效空间相关系数图

由图 3 可见,相比于图 1,站点间的空间相关关系更加简洁清楚,对于空间距离权重也有所反映,但不像图 2 那样有强假设的边界。从预测准确率来看,此方法也较前两种更优,见表 2。在模型预测效果评估中,采用检验回归模型较普遍的两个评价指标:均方根误差 (RMSE) 和绝对误差 (MAE),具体表达式为:

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2}$$
$$\text{MAE} = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|$$

其中,  $m$  为样本数,  $y_i$  为真实值,  $\hat{y}_i$  为估计值。

表 2 几种条件高斯图模型预测性能对比

预测时效	误差	MS-CGGM		CGGM		带边界的 CGGM	
		降水	气温	降水	气温	降水	气温
6 小时	RMSE	0.063 1	0.099 7	0.283 6	0.323 4	0.328 0	3.128 2
	MAE	0.043 3	0.062 4	0.235 8	0.240 1	0.265 6	2.988 0
24 小时	RMSE	0.112 5	0.163 7	0.209 7	0.332 1	1.493 7	1.320 8
	MAE	0.093 7	0.116 4	0.173 6	0.248 1	1.265 9	1.214 6



续表 2

预测时效	误差	MS-CGGM		CGGM		带边界的 CGGM	
		降水	气温	降水	气温	降水	气温
72 小时	RMSE	0.135 2	0.177 0	0.147 5	0.326 8	1.075 7	1.204 7
	MAE	0.099 8	0.123 8	0.118 3	0.243 9	0.843 6	1.107 6
7 天	RMSE	0.172 0	0.188 0	0.207 7	0.323 7	0.352 9	0.437 7
	MAE	0.130 9	0.132 8	0.164 6	0.240 0	0.274 2	0.520 9

并且,多尺度条件高斯图模型的运行效率有明显提升,初始条件的改善使得收敛速度更快,迭代次数由平均 496 次减少到平均 321 次,运行时间明显变短。

3 结束语

根据预测时效的不同,提出了多尺度空间相关模型,在条件高斯图模型的基础上进行改进,引入相似度矩阵,对协方差矩阵进行迭代前的先验约束,减少迭代次数,提高效率。实验结果表明,多尺度空间相关模型能够快速学习全球气象观测站的空间相关性,空间相关根据预测时效不同而表现出差异,预测效率和准确率都有所提升。

参考文献:

[1] KOLLER D,FRIEDMAN N. Probabilistic graphical models; principles and techniques – adaptive computation and machine learning[M]. London:MIT Press,2009.

[2] 张宏毅,王立威,陈瑜希. 概率图模型研究进展综述[J]. 软件学报,2013,24(11):2476–2497.

[3] EDWARDS D. Introduction to graphical modelling[M]. New York:Springer,2000:235–254.

[4] 周志华. 机器学习:Machine learning[M]. 北京:清华大学出版社,2016.

[5] 刘建伟,黎海恩,罗雄麟. 概率图模型学习技术研究进展[J]. 自动化学报,2014,40(6):1025–1044.

[6] 刘建伟,黎海恩,罗雄麟. 概率图模型表示理论[J]. 计算机科学,2014,41(9):1–17.

[7] 杜 博,陈 勇,史瑞芝. 一种基于高斯马尔可夫随机场的异常目标探测方法[J]. 测绘科学,2010,35(6):180–182.

[8] 标 本,梁恺彬,管一弘. 高斯马尔可夫随机场的人脑 MR 图像分割方法[J]. 计算机技术与发展,2017,27(7):180–184.

[9] 王文辉,冯前进,刘 磊,等. 基于类自适应高斯-马尔可夫随机场模型和 EM 算法的 MR 图像分割[J]. 中国图象图形学报,2018,13(3):488–493.

[10] LAURITZEN S. Graphical models[M]. [ s. l. ]; Clarendon Press,1996.

[11] FRIEDMAN J,HASTIE T,TIBSHIRANI R. Sparse inverse covariance estimation with the graphical lasso[J]. Biostatistics,2008,9(3):432–441.

[12] LI Hongzhe,GUI Jiang. Gradient directed regularization for sparse Gaussian concentration graphs,with applications to inference of genetic networks[J]. Biostatistics,2006,7(2):302–317.

[13] LU Zhaosong. Smooth optimization approach for covariance selection[J]. SIAM Journal on Optimization,2009,19(4):1807–1827.

[14] EL GHAOU L,BANERJEE O,D’ASPREMONT A. Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data[M]. Cambridge:MIT Press,2008.

[15] SCHEINBERG K,MA Shiqian,GOLDFARB D. Sparse inverse covariance selection via alternating linearization methods[C]//International conference on neural information processing systems. Sydney:Curran Associates Inc. ,2010.

[16] HSIEH C J,SUSTIK M A,DHILLON I S,et al. Sparse inverse covariance matrix estimation using quadratic approximation[C]//International conference on neural information processing systems. Granada,Spain;Curran Associates Inc. ,2011:2330–2338.

[17] WYTOCK M,KOLTER J Z. Sparse Gaussian conditional random fields: algorithms, theory, and application to energy forecasting[C]//Proceedings of the 30th international conference on machine learning. Atlanta:[ s. n. ],2013:1265–1273.