

氏 名	Andrade Silva Daniel Georg
学位(専攻分野)	博士(統計科学)
学位記番号	総研大甲第 2112 号
学位授与の日付	2019 年 9 月 27 日
学位授与の要件	複合科学研究科 統計科学専攻 学位規則第6条第1項該当
学位論文題目	Bayesian Model Selection under Noise - From Statistical to Practical Significance
論文審査委員	主 査 准教授 日野 英逸 教授 栗木 哲 教授 福水 健次 教授 武田 朗子 東京大学大学院情報理工学系研究科 准教授 山田 誠 京都大学大学院情報学研究科

(様式 3)

博士論文の要旨

氏 名 Andrade Silva Daniel Georg

論文題目 Bayesian Model Selection under Noise - From Statistical to Practical Significance

In this thesis, we consider the problem of model selection, and in particular the situation where modeling assumptions regarding exact zero partial correlations or exact zero regression coefficients are violated. For the purpose of getting scientific insights, one is often interested in model selection, i.e. choosing among two or more candidate models. In this thesis we employ the marginal likelihood for Bayesian model selection. The marginal likelihood can identify the most plausible model, or a subset of plausible models, and this way can help the data analyst to gain new insights. One advantage of the marginal likelihood is that it incorporates a model complexity penalty that helps to prefer simpler models over complex models. In general, due to the ease of interpretation, models with low complexity (small dimensional parameter space) are preferred over complex ones.

However, with small noise on the correlation between variables, or small noise on linear regression coefficients, more complex models tend to be selected *independent of the effect size*. This problem is especially pronounced by large sample sizes. In this thesis, we address this problem by carefully designing priors that absorb overly complex models with hyper-parameters that control the desirable effect size. In particular, we address the problem of clustering variables in the Gaussian graphical model (Chapter 2) and variable selection in linear regression (Chapter 3) under such small negligible noise.

In Chapter 2, we address the problem of clustering variables in the Gaussian graphical model. Variable clustering is important for explanatory analysis. However, only few dedicated methods for variable clustering with the Gaussian graphical model have been proposed. Even more severe, small insignificant partial correlations due to noise can dramatically change the clustering result when evaluating for example with the Bayesian Information Criteria (BIC). We address this issue by proposing a Bayesian model that accounts for negligible small, but not necessarily zero, partial correlations. To address the intractable calculation of the marginal likelihood, we propose two solutions: one based on a variational approximation, and another based on Markov Chain Monte Carlo (MCMC).

Our variational approximation is based on a convex optimization problem for finding the maximum a-posterior (MAP) estimate and a low-dimensional non-convex optimization problem for identifying the variance around the MAP. Although, the

former is a convex optimization problem, the high dimension and positive-definite constraint on the precision matrix are challenging for standard convex solvers. Therefore, we adapt a recently proposed 3-block alternating direction method of multipliers (ADMM) to our problem, which proves to be numerically stable and sufficiently precise.

Experiments on simulated data shows that, in the no-noise setting, our proposed performs similar accurate to BIC in identifying the correct clusters, but is considerably more accurate when there are noisy partial correlations. Furthermore, on real data the proposed method provides clustering results that are intuitively sensible, which is not always the case when using BIC and its extensions. Experimentally, we also confirm that the variational approximation is considerably faster than MCMC while leading to similar accurate model selections.

In Chapter 3, we extend some of the ideas from Chapter 2 to variable selection under noise in linear regression. Sparseness of the regression coefficient vector is often a desirable property, since, among other benefits, sparseness improves interpretability. Therefore, in practice, we may want to trade in a small reduction in prediction accuracy for an increase in sparseness. The work in (Chipman et al., 2001) introduces two spike-and-slab priors that can potentially handle such a trade-off between prediction accuracy and sparseness. For that purpose, they introduce a threshold δ on the magnitude of each regression coefficient. Their first spike-and-slab model couples the response variance with the variance on the regression coefficients leading to a closed-form analytic solution. However, as a result, their method is sensitive to the prior setting of the response variance and cannot guarantee anymore that the true model is selected. Their second spike-and-slab prior model solves the latter issue, but at the cost of losing conjugacy. Another subtle issue common to these spike-and-slab priors is that they lead to inconsistent Bayes factors. Due to the fact that their spike-and-slab priors have full support, the Bayes factors of any two models is bounded in probability for increasingly large sample sizes. This is an undesirable property for Bayesian hypotheses testing.

Our proposed model decouples the response noise prior variance from the regression coefficients' prior variance, and thus makes the threshold parameter δ more meaningful than previous work. For example, δ can be set such that the Mean-Squared Error (MSE) of the prediction is only little influenced by ignoring covariates with coefficients' magnitude smaller than δ . In case where the specification of δ is difficult, we show that automatic selection of δ via the estimation of MSE can be a viable choice. Furthermore, by using disjunct support priors, our method guarantees consistent Bayes factors in the sense that the ratio of the true model's marginal likelihood to any other models' marginal likelihood converges to infinity for increasingly large sample sizes. Due to the non-conjugacy of the priors proposed by our method, estimating the marginal likelihood explicitly is computationally infeasible. Instead, we propose to

estimate all model probabilities by introducing a latent variable indicator vector and sampling from its posterior distribution with an efficient Gibbs sampler.

On several synthetic data sets, we evaluate our proposed method in terms of the ability to identify the true model. Here, we define the true model as the one correctly separating all variables into two sets S and C , where S contains all variables that have non-negligible regression coefficients, and C contains all remaining variables.

We compare our method to the spike-and-slab priors as in EMVS (Ročková and George, 2014), GibbsBvs (Bayarri et al., 2012), thresholding the mean regression coefficient vector of an horseshoe prior (Carvalho et al., 2010), and (penalized) maximum likelihood estimation combined with stability selection, AIC, BIC, and its extensions. In various settings: with/without noise and low/high dimensions the proposed method leads to consistently good model selection performance, which was not the case for any other baseline method.

Finally, we evaluated our method also on three real data sets. Concerning the number of selected variables of our proposed method and all previous methods, we observe a similar behavior as for the synthetic data set. Furthermore, for $\delta = 0$, our proposed method seems to roughly agree with various previous methods, while the inspection of the results for $\delta = 0.5$, allows us to draw conclusions about the practical relevance of some of the selected variables.

博士論文審査結果

Name in Full
氏名 Andrade Silva Daniel Georg

論文題目 Bayesian Model Selection under Noise - From Statistical to Practical Significance

(博士論文審査結果) [2019年 7月 25日 実施]

午前10時から約2時間半にわたり、Andrade氏の博士論文予備審査を、本人および5名の委員全員の出席のもとに行った。出願者による1時間にわたる公開発表による概要説明と質疑応答、さらに約1時間の審査員のみによる審査と口述による試験を踏まえた上での審議の結果、以下に述べる理由により審査委員会は全員一致で提出された博士出願論文は学位授与の水準に達していると判断した。

[論文の概要]

博士出願論文は、ノイズが存在する場合のベイズ的モデル選択に関して論じたもので、英文で書かれており全4章118頁からなる。

1章は本論文の序章である。ベイズ的モデル選択として周辺尤度最大化の方法が説明され、関連する従来法が説明されている。

2章は、ガウシアングラフィカルモデルによる変数クラスタリングを論じ、周辺尤度最大化に基づいたクラスタリング法を提案している。精度行列に微小なノイズ成分を加えることによって共分散構造をモデル化し、ノイズに対するロバスト性を持つモデルにしている。対数周辺尤度は積分の陽な表示を持たないため、変分ベイズ法に基づく効果的な近似計算法を提案している。まず変分ベイズ法に要するMAP解は自明でないため、これを3ブロック交互方向乗数法によって解き、この解を代入した変分ベイズ法が1次元の最適化問題に分解されることを用いて効率的な近似計算を行っている。また、候補となるクラスタリングの種類を削減するため、スペクトラルクラスタリングによって候補集合を絞る方法を提案している。人工データに用いて提案手法をBICやグラフィカルモデルに基づく他のクラスタリング法と比較することにより、ノイズが存在しない場合は同等の性能を、ノイズが存在する場合には提案手法が優位性を持つことを確認している。また、3種の実データに適用し既存研究と同様の変数を選択することを確認している。

3章では、線形回帰において有効な説明変数を選択する問題に対するベイズ的方法を提案している。無関係な変数の係数にノイズがあるモデルとして、既存研究ではSpike-Slabモデルによる事前分布が用いられているが、従来のSpike-Slabモデルでは選択される変数が多くなりすぎるといった問題点があった。Spike-Slabモデルを改良し、無関係な変数を表現する密度関数と有効な変数の密度関数が疎な台を持つような事前分布を導入することによってこの問題の解決をはかり、Gibbsサンプラーによる事後確率の近似計算法を提案

している。また、従来の Spike-Slab モデルと異なり、ベイズファクターに関する一致性を持つという理論結果を示している。人工データおよび実データを用いて関連する従来法と比較を行ったところ、同程度以上の性能を持つことが示されている。

4 章は論文のまとめである。

[論文の評価]

ベイズ的モデル選択は重要な問題であるが、本論文は特に、従来十分な研究がなされていなかったノイズに対してロバストなモデル選択を対象とし、実行可能な計算量で有効な方法を提案しており、統計科学の博士論文として十分な意義を持つと考える。なお、2章の内容をまとめた論文が査読付き国際学術雑誌 *Statistics and Computing* (Springer 刊) に採択されている。

(試験結果) [2019 年 7 月 25 日実施]

総合研究大学院大学複合科学研究科における課程博士及び修士の学位の学位授与に係る論文審査等の手続き等に関する規程第 10 条に基づいて、口述による試験を実施した。

口述による試験を実施した結果、出願者はその博士論文を中心としてそれに関連がある専門分野及びその基礎となる分野について博士（統計科学）の学位の授与に十分な学識を有するものと判断し、合格と判定した。