Contributed comment on Article by Wade and Ghahramani

Roberto Casarin* and Stefano Tonellato†

Abstract. This article discusses the Wade and Ghahramani's (2017) paper on a new estimator for clustering structures based on the variation of information (VI) metric. The present discussion focuses on the estimation of concentration parameter of the Dirichlet process. In estimating the clustering structure, the concentration parameter is integrated out and the marginal posterior distribution of the random partition is used to evaluate the posterior loss. Here we propose to use the optimal VI for model selection.

MSC 2010 subject classifications: Primary 62G05, 62F15, 60G57, 60G09.

Keywords: Bayesian nonparametrics, Dirichlet process prior, model selection, variation of information criterion.

1 Introduction

The authors are to be congratulated on their excellent intuition, which has culminated in the development of a new Bayesian point estimator for clustering structure which can find applications in many Bayesian nonparametrics studies. Their Bayesian approach to clustering estimation is inspired by the paper of Meilă (2007). The proposed model provides an alternative to the Dahl (2006) method widely used in the Bayesian nonparametric literature.

In the application to the galaxy data we assume the same DP mixture model as in equation (5) of the paper and the same prior setting $\mu_0 = \bar{x}$, $c = 1/2$, $a = 2$ and $b = s^2$. Instead of estimating α we assume the concentration parameter α takes values in the finite set $A = \{\alpha_1, \ldots, \alpha_n\}$ and for each element α_j of this regular grid we evaluate the partition posterior distribution $p(\mathbf{c}|y_{1:N}, \alpha_j)$ given by

$$p(\mathbf{c}|y_{1:N}, \alpha_j) \propto \frac{\Gamma(\alpha_j)}{\Gamma(\alpha_j + N)} \alpha_j^{k_N} \prod_{j=1}^{k_N} \Gamma(n_{n_j}) m(\mathbf{y}_j),$$

where $m(\mathbf{y}_j)$ is the marginal likelihood of the observations in the j-th partition. For each value of $\alpha \in A$ we run the Gibbs sampler as in the algorithm 8 of Neal (2000) and find the optimal value of the VI criterion at α (VIC_α) as

$$\mathrm{VIC}_\alpha = \min_{\hat{\mathbf{c}}} \int L(\mathbf{c}, \hat{\mathbf{c}}) p(\mathbf{c}|y_{1:N}, \alpha) d\mathbf{c}.$$

*Department of Economics, University Ca' Foscari of Venice, Cannaregio 873, 30121, Venezia, Italy, r.casarin@unive.it

†Department of Economics, University Ca' Foscari of Venice, Cannaregio 873, 30121, Venezia, Italy, stone@unive.it

c	VIC_α			VI
	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 1.5$	
1/2	0.61	0.72	0.83	0.74
1/10	0.23	0.32	0.41	0.29

Table 1: Optimal VIC_α for $\alpha \in \{0.5, 1, 1.5\}$ and different values of c.

The VIC_α obtained are given in Table 1 for $\alpha \in \{0.5, 1, 1.5\}$. The optimal VI value with α integrated out, using a $\mathcal{G}a(1, 1)$ prior, is reported in the last column. The minimumVIC_α is attained for $\alpha = 0.5$ and it is always smaller than the integrated VI. The result suggests the VIC_α is favoring smaller values of the concentration parameter.

2 Conclusion

In their paper, the authors sketch a number of possible extensions. We would suggest as further research line also the combination of posterior clustering probabilities obtained from the different models. It is clear that this is an exciting and stimulating work. We are therefore very pleased to be able to propose the vote of thanks to the authors for their work.

References

Dahl, D. B. (2006). “Model-based clustering for expression data via a Dirichlet process mixture model.” In Do, K.-A., P. Muller, P., and Vannucci, M. (eds.), *Bayesian Inference for Gene Expression and Proteomics*, 201–218. Cambridge University Press. MR2706330. 604

Meilă, M. (2007). “Comparing clusterings—an information based distance.” *Journal of Multivariate Analysis*, 98(5): 873–895. MR2325412. doi: `https://doi.org/10.1016/j.jmva.2006.11.013`. 604

Neal, R.M. (2000). “Markov Chain sampling methods for Dirichlet process mixture models.” *Journal of Computational and Graphical Statistics*, 9:249–265. 604

Wade, S. and Ghahramani, Z. (2017). “Bayesian Cluster Analysis: Point Estimation and Credible Balls.” *Bayesian Analysis*. 604

Acknowledgments

This research used the multiprocessor cluster system at University Ca’ Foscari of Venice (SCSCF).