

**Report on manuscript COST-D-15-00148: “Clustering over user features and latent behavioral functions with dual.view mixture models.”**

*Octobre 2015.*

---

The paper discusses the use of a dual mixture of Dirichlet process (MDP) model to cluster users who share similar features and similar latent functions, the latter representing individual behavior observed only through their output.

General comments:

- The setting proposed by the authors in section 5 include a gamma distribution for the Dirichlet total mass parameter  $\alpha$ . However, their exploration of properties and performance through the synthetic data is carried out with a fixed non-reported value (page 9 lines 17-18). This is relevant as the number of clusters is given implicitly by the number of distinct labels in the MDP and its prior distribution is highly dependent on the parameter  $\alpha$ , it is leptokurtic and when fixed it gives strong information on the posterior number of clusters [1].
- If the number of components in the mixture is considered fixed, the mixture has Gaussian components and the data are simulated from the baseline distributions. This scenario could be different if the clusters came from a non-Gaussian distribution. The mixture would use more components to model the density properly, overestimating the number of clusters.
- Examples reported in detail are those with non-overlapping clusters, from my perspective it is really possible to assess the performance of a clustering method when clusters overlap. The classified data in the overlapping scenarios is not exhibited and it would be useful to display it as well.

Given the three points described above, I believe there may be a misleading comparison among the models, called dual-fixed, dual-DP and single.

Specific comments:

- Section 5.1. The footnote gives the expectation for a random variable  $x \sim \mathcal{G}(v, w)$ . It is incorrect, depending on the parametrization used it should be  $\mathbb{E}(x) = vw$  or  $\mathbb{E}(x) = v/w$

## References

- [1] LIJOI, A., MENA, R.H. AND PRÜNSTER, I. "Controlling the reinforcement in Bayesian non-parametric mixture models", *J. R. Statist. Soc. B* **69**, 715-740, 2007.