

Clustering over user features and latent
behavioral functions with dual-view mixture
models
-Reviews-

Dear Editors,

...

List of detailed responses:

(Q: editor comments. A: authors replies.)

Editor-in-Chief

Q1: Let me emphasize what the AE already mentioned: "Maybe modify the figures to not only differentiate groups by color, but also with different plot symbols so that they can also be distinguished when printed in gray scale."

Our print issue by default gets published in grayscale (unless you want to pay for color figures). So many of the differently colored points in your figures will become almost indistinguishable for readers of our print issue.

A1: ...

Q2: Please minimize the use of footnotes and add the text back to the main text, e.g., in parentheses.

A2: We have integrated into the text the footnotes of page 8 (about MLEs), page 11 (Geweke's test), page 14 (users/thread ratio) and page 16 (citation of R package mclust).

Q3: In footnotes 3 and 6, you refer to the coda and mclust R packages. Formally cite R and R packages you use in the main text and add to the References. Look at a recently published article from our journal for reference.

A3: We have added references to both packages.

Q4: Place table captions above the table.

A4: We have placed caption above the table (page 11)

Q5: Please check the completeness of your citations, in particular with

respect to missing page numbers and missing editor names for conference proceedings.

A5: We have added missing pages and editors. Now all conference proceedings have their corresponding pages and editors.

Associate Editor

Q1: Please consider changing behaviors view to behavior view or behavioral view.

A1: We replaced *behaviors view* by *behavior view*. To keep consistency, we also replaced *features view* by *feature view*.

Q2: Maybe try to emphasize more already at the early beginning of the manuscript that the feature and the behavioral cluster structure are the same and maybe explain why and when they could be assumed to be aligned; otherwise the improvement in the cold-start problem is unclear.

A2: We have added more emphasis on that in the first and second paragraphs of the Introduction. In the second paragraph, we added some more explanation to the scenario of the cold-start problem.

Q3: The variables used by Cheng et al. (2014) are unclear.

A3: We have specified the type of used variables (page 3).

Q4: Bonilla et al (2010) and Abbasnejad et al (2013) seem to not use dual-view clustering. This should be more explicitly stated and it should be made clearer why they are mentioned, i.e., their relationship to social network clustering.

A4: The main difference is actually that they do not use dual-view (i.e.: they do not consider user features). However, they are similar in that users in the same cluster are assumed to have similar behavioral functions. We have clarified this in the paper (page 3, before Section 2)

Q5: One could state that the generalization to infinite number of clusters leads to a non-parametric model.

A5: We had mentioned this fact at the end of the Introduction (page 3). Besides, we have explicitly reminded this again in Section 2.3. We also suggest adding “non-parametric” to the title of the paper.

Q6: Below Equation (1) please correct p_k to p_K .

A6: We have fixed the subscript from p_k to p_K .

Q7: The statement “In this case, every component is associated to a cluster.” should be toned down as there also exist applications of mixtures where several components are used to form a cluster.

A7: We have changed it by *is often*.

Q8: Use consistent bold or non-bold face for a_u , α and index u .

A8: We have changed a_u by \mathbf{a}_u and α by α in equation (3).

Q9: I am not sure if Equation (7) is correct. It gives the impression that this holds for all k where $n_k = 0$, i.e., an infinite amount of k . But rather I would assume that the combined weight of all empty components should be proportional to α . This comment also holds for Equation 24. Also Equations 25 and 26 could be modified to avoid the restriction of k within the probabilities.

A9: The AE is right. We have fixed the notation in equations (7), (24), (25) and (26).

Q10: It is unclear why the set $U = \{u_1, \dots, u_U\}$ is introduced, i.e., why referring to $1, \dots, U$ is not sufficient. For $\mathbf{a}_u = (\mathbf{a}_1, \dots, \mathbf{a}_D)$ the question arises if a subindex u should also be used for a_i , i.e., a_{ui} .

A10: We have removed the sets \mathcal{U} and \mathcal{T} since they are not used. We have added the subindex u to each element of vector \mathbf{a}_u (page 6)

Q11: Below Equation 17 a different notation is used for the precision.

A11: We have fixed this typo.

Q12: It would be good to provide some justification, e.g., a reference, for the use of the ML estimators for the priors. One might also explain why the use of these estimates is sufficiently flexible if used as priors for the mixture components.

A12: We have no references using this approach. We used MLE to replicate the idea of center the hyperparameters at the data, applied in the feature view. We have added a comment on that in the text (page 9).

Q13: One might explain why choosing $m = 3$ was assumed to be sufficient.

A13: We added some more discussion before Section 4.1 and referred the reader to Neal 2000 for a more systematic discussion.

Q14: Add a reference for the iris dataset. Note also that Anderson (1935)'s iris dataset would seem to consist of more observations than used in the example and more variables.

A14: We have added a reference to the Anderson (1935), and specified that we use the version available in R as well as the three used features. We have also clarified that we use dropped one of the variables and used a random subset of 50 observations so that the clustering really needs the behavior data to improve the results. (page 16)

Q15: Maybe explain why different initialization schemes are used for the two experiments and the iris dataset.

A15: Different initialization schemes are used to show that the inference for this model is not too sensitive to different initializations, though the k-means strategy would be in general a smarter choice than initializing with a single cluster. We added some discussion at the end of Section 5.1 (page 11).

Q16: Maybe modify the figures to not only differentiate groups by color,

but also with different plot symbols so that they can also be distinguished when printed in gray scale.

A16: ...

Q17: Regarding the additional analysis using the iris dataset it would be good to better motivate it. Please specify which observations are included, which variables are included and more clearly explain why it was complemented with artificial data and not a real dataset was used which already contains all required features for applying the proposed model. When evaluating the results it would be good to point out that the class labels are used as ground truth for evaluation while, however, it is unclear in cluster analysis if recovering true class labels is necessarily the aim.

A17: We have specified the three variables used and explained that we use a random subsample of 50 observations so that the clustering really needs the behavior data to improve the results. (page 17) We have also added some discussion on why we chose the iris dataset and not real forum data (before Section 5.1, page 10).

Q18: Maybe add round parentheses around vectors, e.g., π and α at the beginning of Appendix A.

A18: We have added round parentheses around these vectors.