

Cluster Analysis of the Rat Olfactory Bulb Activity in Response to Different Odorants

M. Falasconi^{a*}, A. Gutierrez^{b,c}, B. Auffarth^{b,c}, G. Sberveglieri^a, S. Marco^{b,c}

^a Dept. of Chemistry and Physics, Univ. of Brescia & SENSOR Lab CNR-INFM, Via Valotti 9, 25133 Brescia, Italy;

^b Departament d'Electrònica, Universitat de Barcelona, Martí i Franquès, 1, 08028 Barcelona, Spain; ^c Artificial Olfaction Group, Inst. for Bioengineering of Catalonia (IBEC), Baldiri i Rexach 13, 08028-Barcelona, Spain

* Corresponding author: Tel: +39 030 3715709; Fax: +39 030 2091271; email: matteo.falasconi@ing.unibs.it

Abstract. With the goal of deepen in the understanding of coding of chemical information in the olfactory system, a large data set consisting of rat's olfactory bulb activity values in response to several different volatile compounds has been analyzed by fuzzy c-means clustering methods. Clustering should help to discover groups of glomeruli that are similarly activated according to their response profiles across the odorants. To investigate the significance of the achieved fuzzy partitions we developed and applied a novel validity approach based on cluster stability. Our results show certain level of glomerular clustering in the olfactory bulb and indicate that exist a main chemo-topic subdivision of the glomerular layer in few macro-area which are rather specific to particular functional groups of the volatile molecules.

Keywords: olfactory bulb, 2-deoxyglucose mapping, olfactory coding, cluster analysis, cluster validity

PACS: 07.05.Kf, 07.05.Pj, 07.05.Rm, 29.85.Fj, 87.19.lt, 87.85.Ng, 89.75.Fb

INTRODUCTION

Odour stimulation of olfactory sensory cells in the mammalian nose involves odour molecules interacting with Olfactory-Receptor Neurons (ORN) located on the cilia of the Olfactory Epithelium (OE). It is known [1] that ORN expressing the same binding protein project onto the same roughly spherical synaptic bundles (glomeruli) of the Olfactory Bulb (OB). As a consequence, the activation of the glomerular layer is a bidimensional representation of the chemical features of the odorants. What it is not totally understood to this date is what chemical features are binded by the ORNs and in turn activate the different glomeruli. It is hypothesized that the ORN detect chemical features as functional groups or carbon chain length.

An extensive experimental study on the OB activity in response to different volatile organic compounds (VOCs) has been conducted in mice by Leon et al. [2] by mapping the uptake of ¹⁴C radiolabeled 2-deoxyglucose (2DG). In about ten years of activity Leon and coworkers collected a database of OB responses to more than 300 different VOCs.

These studies show that odour stimuli in the rat produce spatial patterns of activity in the glomerular layer that are overlapping but that vary for different

odours. The analysis of such activity patterns is a challenging problem. Odour patterns are complex, highly irregular, noisy, and may contain missing values due to the experimental procedure. For these reasons, automated exploratory data analysis techniques are required for objectively investigating the data structure.

In this work we used *fuzzy c-means* (FCM) cluster analysis to discover groups of pixels that have similar response patterns across the entire set of VOCs. In order to estimate the significance of the returned results we have developed an original validity criterion for fuzzy partitions [3] based on the evaluation of partition stability under bootstrap resampling of the data which allows to estimate the best partition of data.

The main objective is a deepen comprehension of the olfactory coding, i.e.: understanding whether and how glomeruli are clustered together according to their response profiles across the odorants and interpreting the retrieved valid clusters in terms of chemical properties of the VOCs.

EXPERIMENTAL AND METHODS

The available data (<http://leonsrver.bio.uci.edu/>) consist of 470 activity maps of the OB glomeruli for

different VOCs obtained by mapping the uptake of ^{14}C -2DG radiolabel over the entire OB layer. Such method provides a digital image (80x44 pixels) that represents a bidimensional map (ventral centered) of glomerular activity. Units in each matrix are standardized to z-scores relative to the mean and standard deviation of values across that matrix. Each map was vectorized and all background pixels (with zero variance) and missing values were removed leading to a final set of 1780 pixels.

The original database comprises measurements of the same chemical at different concentrations, thus the actual number of different VOCs is about 300. To eliminate the concentration effects and focus only on the chemical features of the molecule, the measurements of the same chemical at different concentrations were averaged together and replaced with a single map. Both data sets (D1=1780x470; D2=1780x308) were then analyzed.

Cluster analysis was performed by the FCM algorithm which outputs fuzzy partitions with a pre-specified number of clusters. We used a novel paradigm for fuzzy partitions validity based on the concept of partition stability under data perturbation. Data can be perturbed in several ways, here we repeatedly apply bootstrap sampling on pixels, so we named the method fBPSE (Bootstrap Partition Stability Estimation, where f stands for fuzzy) [3].

The fBPSE estimates and minimizes, over a series of partitions with increasing number of clusters K , the partition variability:

$$V(K) = \frac{1}{\binom{B}{2}} \sum_{i=1}^B \sum_{j>i}^{B-1} d(U_K(Y_i), U_K(Y_j)) \quad (1)$$

where B is the number of bootstrap samples ($B=20$). $V(K)$ is formally equivalent to the empirical bootstrap estimate of the variance, where $d(.,.)$ plays the role of a distance. Actually, $d(.,.)$ is similarity measure between two fuzzy partitions [4], being $U_K(Y_i)$ the membership matrix achieved by running the FCM algorithm on the i -th bootstrap sample Y_i . In our work we adopted the adjusted *fuzzy Rand index* [5].

RESULTS AND DISCUSSION

The mean and the variance of activity values for each pixel were calculated across the 308 different odorants of data set D2 (the results are very similar for D1).

The map of the mean activity (Figure 1) shows a region of high activity in the medial and lateral aspects of the OB while the ventral aspect seems to be almost

inactive. Indeed, it seems that the lateral aspects of the OB are more active independently of the type of VOC under analysis. As pointed out by Leon et al. [2], the ventral aspect responds selectively to odorants with dense hydrocarbon features and with no oxygen-containing functional groups, however these are a minority in our database.

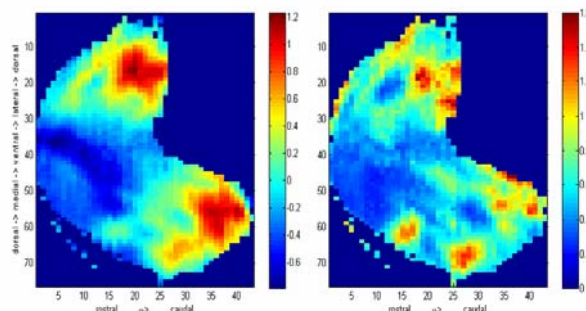


FIGURE 1 Distribution maps of the mean (left) and variance (right) activity values across the entire spectrum of 308 different chemicals of data set D2.

The average variance distribution of pixels (Figure 1) tells a different story. Some selected groups of pixels with higher variance – i.e. higher information content - emerge in the dorso-medial and dorso-lateral regions of the OB whilst the ventral area has no variance. This observation partially confirms the previous one, but also strongly supports the hypothesis that there is a finer clustering structure inherent to the data that should be recovered by partitions with a higher K value.

We then performed FCM clustering and fBPSE validity from $K=2$ up to $K=40$. The difference of activity between the medio-lateral and ventral aspects is the main source of variance in the data set; this was confirmed by Principal Component Analysis (PCA) since the distribution map of PC1 is very similar to the left picture in Figure 1 (data not shown). Therefore, being the FCM method based on the intra-cluster variance minimization, the partition $K=2$ returns two clusters: the first cluster comprises the pixels in the medio-lateral aspects, the second one coincides with the ventral area.

Moreover all the achieved partitions show clusters of pixels that are paired in the medial and lateral aspects of the OB (see for example the clusters in Figure 3). This was an important consistency check for clustering because it is known that glomeruli are co-activated in the medio-lateral regions of the bulb.

By visually exploring the achieved partitions we observed a good qualitative agreement between certain clusters and the chemo-topic response model developed by Leon and coworkers. This was another argument in favour of our clustering procedure.

Finally, we noted that some clusters, typically those matching modules with high specificity towards particular VOCs (e.g. carboxylic acids), were emerging in low K partitions and have a long life-time, i.e. they survive almost unaltered for several K values before splitting in smaller clusters. This indicates the presence of individually significant clusters, and then opens new perspectives for further validity studies of single clusters..

The cluster validity study aimed to determine the best number of clusters. However we have not found a clear evidence for this optimum value and our findings do not support the results previously obtained by Leon et al.

We found out that the best partition, i.e. most stable according to the fBPSE, was at K=3 (Figure 2) with a rather similar variability value for K=2.

Similar results were obtained on the two data sets D1 and D2, this indicates that the presence of measurements at different concentrations has negligible effects on the FCM outcomes and on the cluster stability.

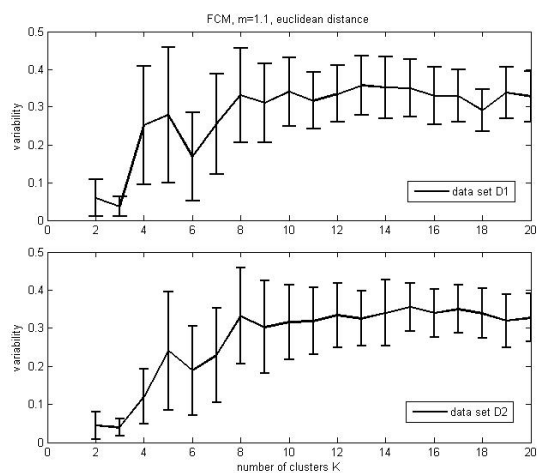


FIGURE 2 Cluster validity results by fBPSE on the data sets D1 and D2. The error bars report the standard deviation of $V(K)$ calculated value over the 190 pairs of compared partitions.

Figure 3 shows the partition K=3 projected on the map. On the left figure we reported the crisp partition obtained by converting the fuzzy outcome according to the maximum membership rule (pixel is assigned to the cluster having maximum membership). To give an insight of the assignment of pixels belonging to the fuzzy cluster borders we showed on the right figure the maximum membership value of each pixel (the brown regions represent the core pixels for which the membership is equal to one).

The high stability value seems associated with the big difference in the average activity between the

lateral and ventral aspects of the OB. However, the partition K=3 shows a further split in the dorso-lateral part of the OB (clusters #1 and #2 in Figure 3) that may indicate an important difference between those areas.

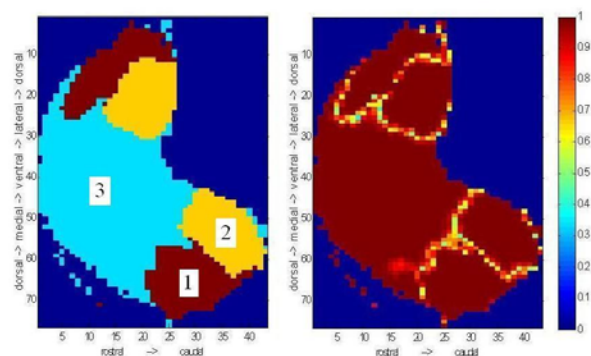


FIGURE 3 Best partition of data in K=3 clusters according to the fBPSE criterion: crisp partition corresponding to the fuzzy clustering (left), maximum value of the fuzzy membership of each pixel (right).

Our findings seem to indicate the existence of a main subdivision of the glomerular activity in three macro-areas.

Such topological representation of the glomerular layer in main zones of activity has been already proposed by Mori et al. [6]. According to Mori, glomeruli are parcelled into four zones in the OB and receive inputs from different regions of the olfactory epithelium. Zone I glomeruli are localized in the rostral-dorsal portion of the OB, while Zones II, III, and IV glomeruli are distributed progressively more ventrally and caudally. The hypothesis is that primary chemical features, such as functional groups, are responsible for the topological organization of glomeruli in these main zones. Secondary features of odorants like carbon chain length are represented by local arrangement of activated glomeruli within each zone.

The existence of such hierarchical structure may explain why we found out a best subdivision of data in three groups, which correspond approximately with the primary activity zones. A preliminary hypothesis is that Zone I corresponds to cluster #1, Zone II to cluster #2, while Zone III-IV are merged together in cluster #3. Why we did not find a subdivision in four clusters remains unexplained. A reasonable hypothesis is that this depends on the available data since many of the measured VOCs do not elicit glomeruli in the ventral region of the OB.

By investigating within each cluster the activation of groups of odorants sharing well defined chemical properties (Figure 4) we noted that: cluster #1 is

mainly activated by ketones (including small aliphatic with less than 8 carbons) and aromatic compounds; in cluster #2 esters dominate; cluster #3 is mainly active for carboxylic acids and aldehydes (when looking at finer partitions, acids form a clear cluster in the dorsal aspect while aldehydes activate the medial region). This seems consistent with previously published works [2, 6].

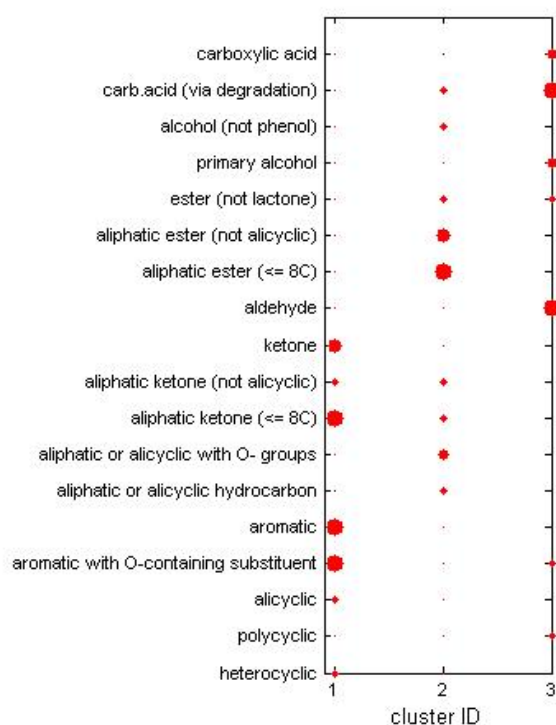


FIGURE 4 For each cluster are reported only the groups of odorants with the highest activation values (cluster ID refers to Figure 3). The bubble size is proportional to the average activity value across the group of odorants sharing the given chemical property.

The intrinsic limitations of our cluster validity method may provide another explanation why we did not recover the modular structure proposed by Leon et al. Both the clustering and the validity procedures are very challenging due to the high dimensionality of the data. For this reason we compared the results obtained in cluster analysis validity with all dimensions to those obtained reducing the dimensionality (taking the first few PCs) of feature space. The outcome was very similar to that achieved with the complete feature set. We have also investigated whether the fBPSE method is biased towards low K values, at present we did not find any evidence for that by testing the method on synthetic data sets [3].

Finally, it can be argued that the current chemo-topic response model was built by Leon's group through a systematic investigation of the glomerular

activity for selected VOCs. On the contrary, in this work we are clustering the glomeruli by taking all the chemicals at the same time. The model construction was more similar to a two-way clustering procedure, in which the pixels of the OB map are iteratively clustered by taking into account only subsets of chemical features. Therefore, further advances in this direction can be made by using two-way clustering methods.

CONCLUSIONS

In this work we applied fuzzy cluster analysis techniques to investigate the response patterns of rat's OB activity towards different volatile compounds aiming to understand the chemotopic organization of the OB. We also approached the challenging problem of validating the achieved partitions by using an original criterion based on fuzzy partition stability.

Our results suggest the existence of certain level of glomerular clustering in the olfactory bulb based on the activity behaviour across the measured VOCs. The results also indicate that exist a main chemo-topic subdivision of the glomerular layer in macro-area (zones) which are active for chemicals differing for the functional group. Further work is currently in progress to interpret the clustering results in terms of chemical features of the odour molecules.

ACKNOWLEDGMENTS

This work was supported by the European Network of Excellence GOSPEL "General Olfaction and Sensing Projects on a European Level" (FP6-IST-2002-507610).

REFERENCES

1. G. M. Shepherd, Smell images and the flavour system in the human brain, *Nature insight* **444** 316-321 (2006).
2. M. Leon, B. A. Johnson, Olfactory coding in the mammalian olfactory bulb, *Brain Research Reviews* **42** 23-32 (2003).
3. M. Falasconi et al., A stability based validity method for fuzzy clustering, *Pattern Recognition* (Submitted, Oct. 2008).
4. C. Borgelt, Resampling for fuzzy clustering, *International Journal of Uncertainty, Fuzziness and Knowledge-Based* **15** 595-614 (2007).
5. R.J.G.B. Campello, A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment, *Pattern Recognition Lett.* **28** 833-841 (2007).
6. K. Mori, Y. K. Takahashi, K. M. Igarashi and M. Yamaguchi, *Physiol Rev* **86** 409-433 (2006).