An Algorithm for Clustering Decision-Making Phenotypes from Behavioural Data

Abraham Nunes (nunes@dal.ca)
Department of Psychiatry & Faculty of Computer Science
Dalhousie University, Halifax, NS, B3H 2E2, Canada

Alexander Rudiuk (rudiuk@dal.ca)
Department of Psychiatry & Faculty of Computer Science
Dalhousie University, Halifax, NS, B3H 4R2, Canada

Thomas Trappenberg (tt@cs.dal.ca)
Faculty of Computer Science
Dalhousie University, Halifax, NS, B3H 4R2, Canada

Abstract

Computational modelling of behavioural data is increasingly used for psychiatric applications. To this end, it is of interest to devise approaches for clustering behavioural phenotypes, where those phenotypes are reinforcement learning models fit to individual subjects. The present study highlights that clustering in the space of individual parameter estimates may not reveal the presence of decision-making phenotype groups. We also introduce a similarity measure that improves cluster detection using synthetic data generated on a task that measures goal-directed and habitual control. Our results motivate further investigation regarding unsupervised cluster detection in computational cognitive modelling of human behavioural data.

Keywords: Computational psychiatry; Clustering

Introduction

The majority of approaches to computational modelling of reinforcement learning in psychiatric populations have focused on parameter estimation and model selection. There is relatively less work done on detecting phenotype clusters, which is the primary concern of this paper. To approach this problem, we must identify reasonable measures of similarity upon which to base clustering procedures.

A simple choice may be distances between reinforcement learning parameters estimated from behavioural data. Although group differences are generally reported in this parameter space, there are drawbacks. There may be many equally likely parameterizations for models generating a given set of behavioural data. Distant parameterizations may also yield similar behavioural data, which would cast doubt upon the validity of a parameter space similarity measure.

We propose a similarity measure that accounts for the likelihood with which subjects' individually optimized parameters can explain each other's data. In terms of ability to inform a successful clustering procedure, our method is compared to a similarity metric based on distances in parameter space. Our paper shows that it may be possible to detect clusters of decision-making phenotypes in a set of behavioural data in
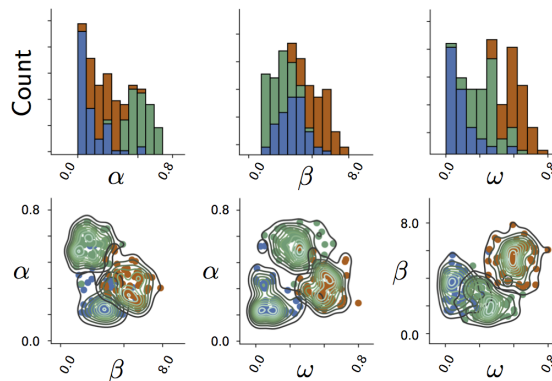

Figure 1: Parameter distributions in three synthetic groups. *Symbols:* learning rate (α), and inverse softmax temperature (β), model-based control weight (ω).

cases where groups do not cluster in the space of parameter estimates.

Methods

We simulated 200 trials of synthetic data generated by 3 groups of 60 agents on a reinforcement learning task previously described by (Kool, Cushman, & Gershman, 2016). This task is primarily used to measure differences in goal-directed and habitual control. Our task differed from that of (Kool et al., 2016) insofar as the random walks governing reward were anticorrelated ($\rho = -1$), and reward magnitudes were continuous values between 0 and 1. Agents were defined as unique parameterizations of a model consisting of softmax action selection (with inverse temperature β), instrumental Rescorla-Wagner learning rule (with learning rate α), and model-based controller implemented with the Bellman equation (Bellman, 1957). Like previous work, we represent the balance of MB and model-free (MF) control as a weighted sum with ω being the proportion of MB control implemented. The three groups' parameterizations are shown in Figure 1.

Using the same model architecture above, we computed subject-wise maximum likelihood estimates (MLE) of parameters using the `fitr` package for Python (Nunes, Rudiuk, &

Trappenberg, 2016). The parameter distance ϕ_{ij}^{Param} between subjects i and j was computed as the cosine distance between θ_i and θ_j.

We then computed a "likelihood distance". For subject i, we began by quantifying the subject-wise difference in likelihood of data generated by subject j, $\mathbf{D}_j$, under the parameters inferred for subject i:

$$\mathbf{x}_i = \left[\log p(\mathbf{D}_j | \theta_i) - \log p(\mathbf{D}_i | \theta_i) \right]^2 \ \forall j \in \{1, \dots n_{subjects}\} \tag{1}$$

The cosine distance between $\mathbf{x}_i$ and $\mathbf{x}_j$ yields an $n_{subject} \times n_{subject}$ matrix $\Phi^{\mathcal{L}}$ henceforth called the likelihood distance matrix.

We subjected the parameter distance matrix Φ^{Param} and likelihood distance matrix $\Phi^{\mathcal{L}}$ to the affinity propagation algorithm (Frey & Dueck, 2007), which we selected in order to determine whether the correct number of clusters in the data could be identified. The primary measure of clustering performance was adjusted mutual information (Vinh, Epps, & Bailey, 2009).

To qualify the difference in clustering performance between parameter and likelihood distances, we plotted for each pair of subjects a point $(\phi_{ij}^{Param}, \phi_{ij}^{\mathcal{L}})$ coloured according to whether they belong to the same true underlying group. If the likelihood distance performs superiorly in this case, greater separation of in- and out-groups should be observed along that axis.

Results

Figure 2 shows the distance matrices Φ^{θ} and $\Phi^{\mathcal{L}}$ respectively, and clustering performance is reported in Table 1. Affinity propagation with likelihood distance out-performed the same algorithm using parameter-space similarity (adjusted mutual information of 0.37 and 0.01, respectively). Clustering upon our similarity measure also resulted in identification of the correct number of clusters in this case (compared to an estimate of 15 clusters based on parameter distance).

Table 1: Clustering performance using parameter distance Φ^{Param} and likelihood distance $\Phi^{\mathcal{L}}$ in the affinity propagation algorithm. *Abbreviations:* adjusted mutual information $\mathbb{I}^{Adj}$

Metric	$n_{clusters}$	Homogeneity	Completeness	$\mathbb{I}^{Adj}$
Φ^{Param}	15	0.11	0.05	0.01
$\Phi^{\mathcal{L}}$	3	0.37	0.37	0.36

Figure 3 shows a comparison in parameter and likelihood distance metrics within and between groups. In-group and out-group distances were greatest along the the likelihood distance, rather than parameter distance.

Discussion

We have shown that unsupervised cluster detection of decision-making phenotypes is possible using behavioural data, but that performance of such clustering methods may
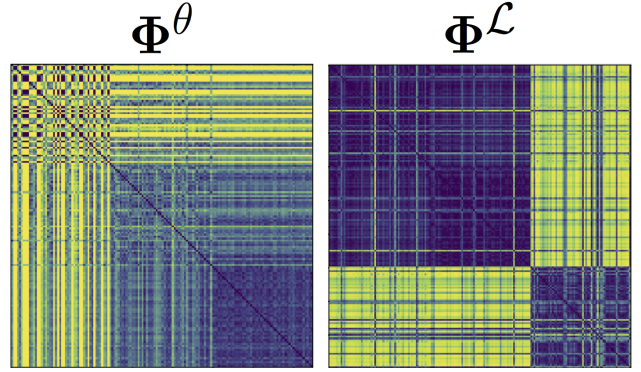

Figure 2: Distance matrices. Φ^{θ} shows parameter distances, and $\Phi^{\mathcal{L}}$ shows likelihood distances. Shorter distances are represented by darker blue colour. Subjects are oriented along the rows and columns in 3 groups (according to the data's generative process), as evinced by the block pattern emergent in the distance matrices.

depend on the similarity measure employed. Our results highlight one important consideration in computational cognitive modelling of human behavioural data. Specifically, that absence of group clustering in parameter space may not reflect the absence of decision-making phenotype clusters in the data. In our case, the inter-subject distances in parameter space were virtually uninformative for cluster identification. However, when the mapping of parameters on to behavioural data was considered, we were able to recover the underlying group structure.

The present study was not designed to show superiority of the likelihood distance measure across all tasks and reinforcement learning models. Rather, we sought to demonstrate the details and feasibility of this approach in a single case. Future work will evaluate the sensitivity of likelihood distance based clustering to variations in task, reinforcement learning model, and group composition. Our work suggests that unsupervised approaches for analysis in computational psychiatry may be of further value.

Acknowledgements

References

Bellman, R. (1957). *Dynamic Programming* (Vol. 70). Princeton, NJ: Princeton University Press.

Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. *Science*, *315*(5814), 972–976.

Kool, W., Cushman, F. A., & Gershman, S. J. (2016). When Does Model-Based Control Pay Off? *PLOS Computational Biology*, *12*(8), e1005090.
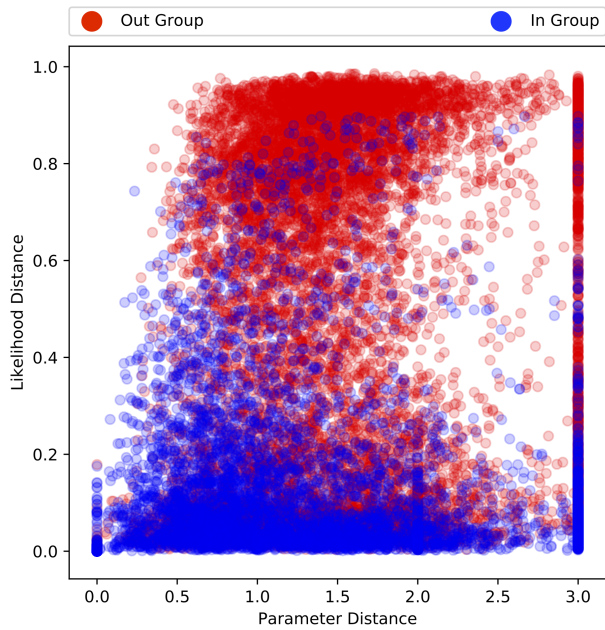
Figure 3: Comparison of group separation using likelihood distance vs. parameter distance. Red points denote subject pairs that are not in the same underlying group. Blue points denote pairs of subjects that are in the same ground truth cluster.

Nunes, A., Rudiuk, A., & Trappenberg, T. P. (2016). *Fitr: A Toolbox for Computational Psychiatry.* Halifax, NS: GitHub. Retrieved from https://github.com/abrahamnunes/fitr doi: 10.5281/zenodo.162263

Vinh, N. X., Epps, J., & Bailey, J. (2009). Information theoretic measures for clusterings comparison. In *Proceedings of the 26th annual international conference on machine learning - icml '09* (pp. 1–8).