Measures of Neural Similarity

Sebastian Bobadilla-Suarez (sebastian.suarez.12@ucl.ac.uk) and Bradley C. Love (b.love@ucl.ac.uk)
University College London and The Alan Turing Institute
Department of Experimental Psychology, 26 Bedford Way
WC1H 0AP London, United Kingdom

Keywords: neural similarity; representational similarity analysis; confusability; task demands; fMRI

Introduction

Detecting similarities between objects is a cornerstone of many cognitive operations, such as categorization and memory retrieval. Although a great deal of effort has been put into understanding the nature of similarity in the behavioral realm, the foundations in neuroscience are less established. For example, powerful analysis techniques that involve comparing the similarities of brain states, such as representational similarity analysis (RSA), typically assume that brain states are similar to the extent that they are Pearson correlated (Kriegeskorte, Mur, & Bandettini, 2008). Each similarity measure brings with it a host of assumptions. For example, Pearson correlation assumes that overall levels of voxel activity are normalized and that each voxel independently contributes to similarity, whereas Minkowski measures assume similarity involves distances in a metrical space instead of vector directions, whereas the Mahalanobis measure expands on both measures by assuming that the distributional pattern of voxel activity is consequential. Which similarity measure best describes the brain's operation? We address open questions such as whether the nature of neural similarity is common across tasks and brain regions.

To that end, we follow the tradition of grounding similarity in confusability; when two things are similar they are easily confused. Complementarily, when two things are dissimilar they are easily discriminated. Confusability of different brain states can be measured with classification procedures such as linear support vector machines. For example, a sparrow may be more likely to be misclassified as a robin than a truck. We then consider which similarity measure best characterizes this confusability data for each brain region and task considered. Because some of the similarity measures we will consider are distance metrics, henceforth we will speak of dissimilarity, which we define as the inverse of similarity.

Methods

We analyzed data from an fMRI study considering the categorization of simple geometric shapes (GS, Mack, Preston & Love, 2013) and a second study that involved judgments of natural images (NI, Bracci & Op de Beeck, 2016).

Trial-by-trial estimates and ROI masking

For both studies, we used a method known as LSS (Least Squares – Separate) trial-by-trial estimation to get a parameter estimate for each individual presentation of each stimulus (Mumford, Turner, Ashby & Poldrack, 2012).

To parcellate the different anatomical regions for each participant we used 110 regions of interest as masks from the Harvard-Oxford cortical and subcortical structural atlases. The masks were transformed from MNI space to each participant's native space.

Classification analysis and ROI selection

We trained a linear SVM classifier with leave-one-out *k*-fold cross validation where *k* was equal to the number of functional runs for each participant and each ROI in each study. An optimization procedure looked for the top *n* voxels (ordered by highest to lowest *F* values from an ANOVA per voxel) **where *n*** maximized linear SVM accuracy for all pairwise classification problems between stimuli in each study with respect to a validation run. At the end of this procedure we rank ordered ROIs with respect to classifier accuracy and selected the union of the top ten regions across both studies for further analysis.

Neural dissimilarity analysis

The goal of this analysis was to compare competing dissimilarity measures for each ROI for each participant. Each dissimilarity measure was evaluated by Spearman correlating its pairwise dissimilarities with the corresponding classifier confusion rates. We computed all pairwise dissimilarities (i.e. for all pairs of stimuli) from the training runs defined in the classification analysis – not including the validation run. An analogous feature (i.e., voxel) selection method was used to maximize Spearman correlation for each dissimilarity measure as was used to tune the SVM classifier.

Dissimilarity measures

We evaluated the following dissimilarity measures: negative dot product, cosine distance, City-block (Manhattan) distance, Euclidean distance, three variants of Minkowski distance (with norms 5, 10 and 50), Chebyshev distance, 1-Spearman correlation, 1-Pearson correlation, three variants of Mahalanobis distance, three variants of Bhattacharya distance, variation of information, and distance correlation. The three variants of Mahalanobis distance and Bhattacharya distance were due to the way the sample covariance matrix was regularized; either no regularization, Ledoit-Wolf shrinkage (r) or diagonal regularization (d). Diagonal regularization was defined as the sample covariance matrix with all the off-diagonal elements set to zero. We only report measures that resulted in non-zero mean Spearman correlation with SVM accuracy.

Results

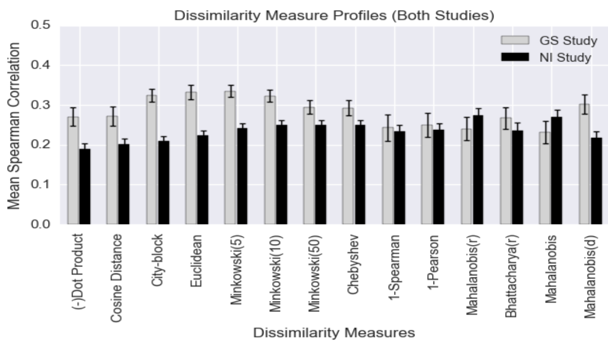

Figure 1: Dissimilarity measure profiles. The mean Spearman correlation for each similarity measure in the GS study (grey bars) and the NI study (black bars) is displayed. The error bars are standard errors of the mean.

The main results are shown in Figure 1. We tested the effect of dissimilarity measures within each study with a mixed effects model. The models contained fixed effects of dissimilarity measure, linear SVM accuracy, participant, and ROI as well as random effects of ROI and participant. For the GS study, the effect of dissimilarity measures was significant, X(2) = 1720.331, p < .001. Similarly for the NI study, the effect of dissimilarity measures was significant, X(2) = 6770.249, p < .001.

The dissimilarity profiles shown in Figure 1 did not correlate between studies, $r(12) = -0.28$, $p = 0.325$, and were significantly different, $t(12) = 4.28$, $p < 0.001$.

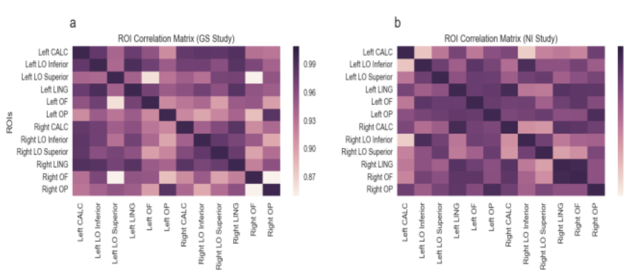

Figure 2: ROI correlation matrices for the (a) GS and (b) NI studies, demonstrating that the performance of dissimilarity measures was Pearson correlated within task. The 12 ROIs (see Methods section) were left and right intracalcarine cortex (CALC), left and right lateral occipital cortex (LO) inferior and superior divisions, left and right lingual gyrus (LING), left and right occipital fusiform gyrus (OF), and left and right occipital pole (OP).

The mean correlation of dissimilarity profiles across the 12 ROIs was 0.95 (SD = 0.034) in the GS study (see panel a of Figure 2) and 0.96 (SD = 0.027) in the NI study (see panel b of Figure 2). Permutation tests (with 10,000 permutations), where the labels of the dissimilarity measures were permuted, showed that the average correlations were significantly different from zero in both studies, $p < 0.001$.

Discussion

We assessed what makes brain states functionally similar by evaluating a wide range of possible dissimilarity measures. Using data from two previous studies, we found 1) measures of dissimilarity differ in how well they gauge relationships between brain states, 2) different brain regions appear to use a common approach to coding state dissimilarity, and 3) the operable measures of dissimilarity vary across studies (i.e., task and stimulus set). These findings suggest that the representation of neural dissimilarity measures morph as a function of task demands or stimuli attributes. Although speculative, the fact that Minkowski measures performed best when stimuli were readily represented in a multidimensional space is suggestive. However, it is possible that differences in how measures of dissimilarity performed across studies were due to differences in data quality, cohort effects, or differences in fMRI equipment.

Acknowledgments

This research was supported by a scholarship from Consejo Nacional de Ciencia y Tecnología (CONACYT) and an enrichment year stipend from The Alan Turing Institute to SBS and by a NIH Grant 1P01HD080679, Leverhulme Trust grant RPG-2014-075 and Wellcome Trust Senior Investigator Award WT106931MA to BCL. We thank all members of the LoveLab for their comments and the fMRI study authors for their data.

References

Bracci, S., & de Beeck, H. O. (2016). Dissociations and associations between shape and category representations in the two visual pathways. *Journal of Neuroscience, 36*(2), 432-444.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience, 2*, 4.

Mack, M. L., Preston, A. R., & Love, B. C. (2013). Decoding the brain's algorithm for categorization from its neural implementation. *Current Biology, 23*(20), 2023-2027.

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage, 59*(3), 2636-2643.