Modeling & Analysis
WE 262

Assessing the reproducibility in sets of Talairach coordinates

Finn Årup Nielsen[1,2], Lars Kai Hansen[2]

[1]Neurobiology Research Unit, Rigshospitalet, Copenhagen, Denmark.
[2]Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark

Abstract

Statements like "this study demonstrates highly consistent findings" or "our results reveal a striking degree of overlap" appear commonly in the literature. Such statements are typically based on informal comparison between activation maps. Computerized methods for comparing activation maps in the form of images exist, e.g., see [1]. Here we propose methods for comparing activation maps when they exist in the form of sets of stereotaxic coordinates. We extented the method developed in connection with information retrieval where a metric was provided for assessing the similarity between the coordinate sets [2]. Our aim is to develop quantitative supports for phrases like "striking degree of overlap" and "highly consistent", i.e., a statistical test for replication, reproducibility or consistency.

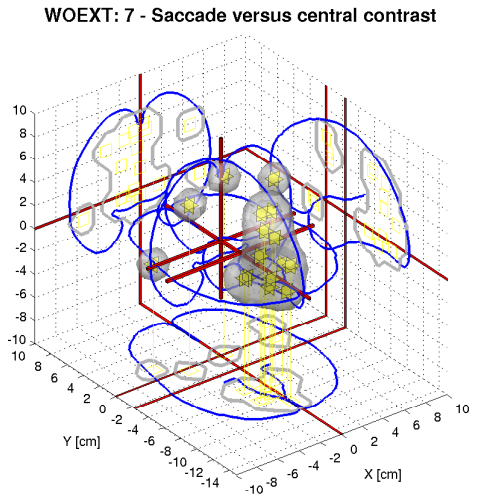
Data: Brede database

Data from the Brede database [3] is used. As the BrainMap database [4] this database is organized in a hierarchical fashion with "papers" on the top level containing one or more "experiments" and with each "experiment" containing one or more "location". The location are associated with Talairach coordinates [5].

The Brede database presently contains 391 experiments from 126 different papers.

Method 1: Volume correlation

We describe two methods. The first method uses a database of "experiments" (sets of Talairach coordinates) to generate a null-distribution for a similarity measure: A distribution is computed for the similarity between all pairs of experiments in the database. When two new experiments are to be assessed for reproducibility their similarity is compared against the distribution of the database. A P-value is generated based on the rank of the similarity. We use a similarity based on voxelization and the cross-correlation coefficient [2], where the voxelization is performed by convolving each location with a Gaussian kernel [6, 7, 8].



Isosurface in example voxelization with an experiment from [9].

Method 2: Minimum distance

Our second method tests whether two coordinates from two different experiments are statistically the same, and the statistic is based on the minimum distance between all pairs of coordinates (xn, xm) in two experiments (the minimum of the minimum)

$$d = \min_{n,m} \left[\sqrt{(\mathbf{x}_n - \mathbf{x}_m)^\mathsf{T}(\mathbf{x}_n - \mathbf{x}_m)} \right] . \qquad (1)$$

To form a P-value a new distance is compared against the distribution found from all pairs of experiments in the database.

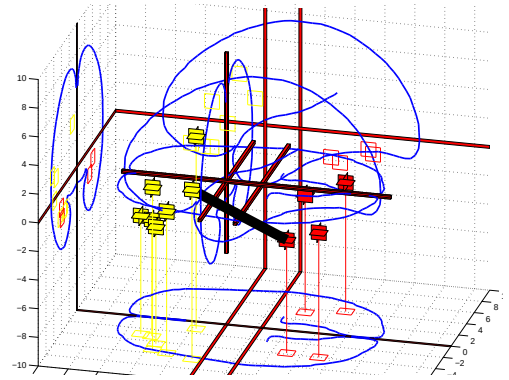
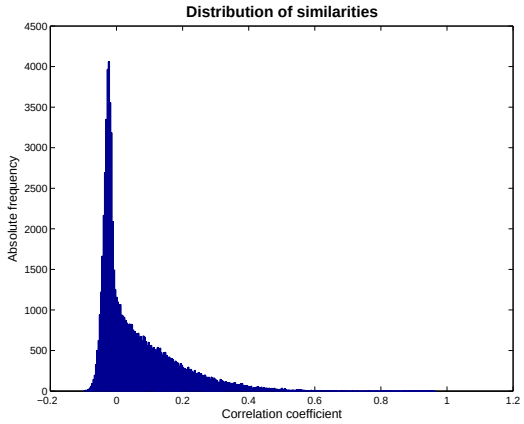


Figure with two experiments a yellow [10] and a red [11] set of locations and with a thick black line indicating the minimum distance.

Results: method 1



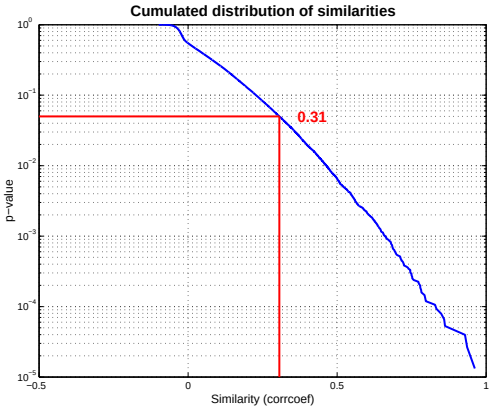
Distribution of similarities between experiments

The above figure shows the sorted similarities from all pairs in the database, excluding those pairs that are from the same paper. The experiments within the same paper is likely to be more correlated the experiments between papers.

Experiments within the same cognitive domain also likely to be more correlated, — but this is presently ignored.

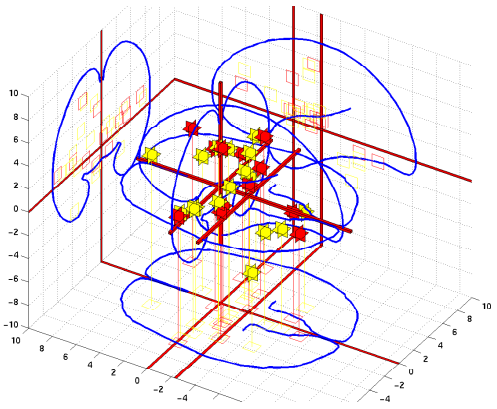
The voxelized experiments have all non-negative elements making the distribution of the cross-correlation coefficient very skewed around zero.

Results: method 1



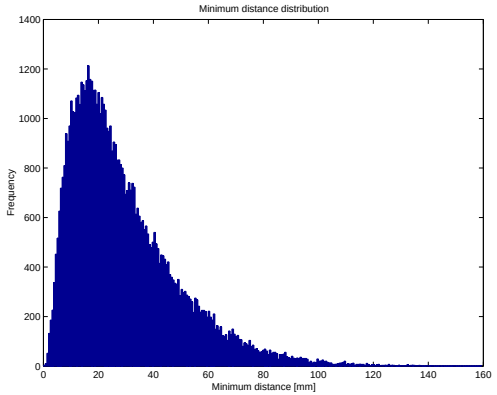
A threshold for $P = 0.05$ appears at a similarity of 0.31.

Example



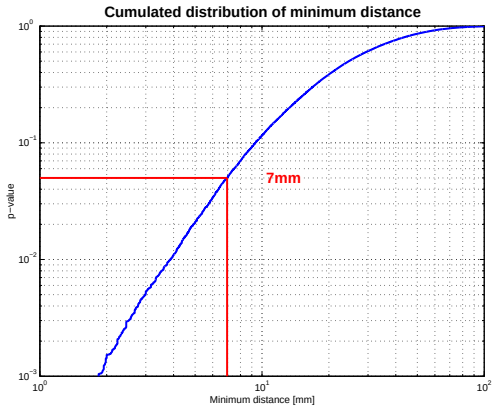
An example on gender differences in pain perception [12] with female as yellow and male as red. Correlation coefficient $c = 0.66$ corresponding to $P = 0.001$

Results: method 2



A histogram of the minimum distance d is shown in figure 2 and the associated d-value for a P-value of 0.05 is $d = 7$mm.

The mean appear at $d_{\text{mean}} = 29$mm and the mode $d_{\text{mode}} \approx 17$mm



Above is the cumulated distribution of the minimum distances.

Example [12]: $d = 5.9$mm corresponding to $P = 0.03$.

Note that the value is a global value indicating similarity on a set level.

Discussion

The distribution of the minimum distance tells us that if we would like to say that two coordinates from two different experiments are the same they should be closer than approximately 7 millimeters, and the similarity distribution indicates that the presently used similarity should be larger than 0.31 before we can accept that an experiment is "reproduced".

The statistics do not model the number of coordinates in each experiment nor their distribution in the brain. One would expect the minimum distance to be smaller if the experiments have many coordinates.

Furthermore, the distribution of the similarity measure changes depending on the type of voxelization and type of similarity measure. The most immediate parameter to model is a strength value such as the z-score or the percent signal change giving more weight to locations with a high value.

Both methods use the same metric in the entire brain.

If one has access to the original volume data it might very well be better to perform the assessment with these data instead of the associated sets of coordinates.

Nevertheless, our method provides a first step for a quantitative reproducibility measure for sets of coordinates.

Availability

The tools for the analysis are available in the Brede neuroinformatics toolbox [13] presently available from http://hendrix.imm.dtu.dk/software-/brede/.

The Brede database is both available in the Brede neuroinformatics toolbox and directly on the Internet: http://hendrix.imm.dtu.dk/services-/jerne/brede/.

Acknowledgment

Finn Årup Nielsen is founded by the Villum Kann Rasmussen Foundation.

References

[1] Lange N, et al. Plurality and resemblance in fMRI data analysis. *NeuroImage*, 1999;10:282–303.

[2] Nielsen FÅ and Hansen LK. Finding related functional neuroimaging volumes. *Artificial Intelligence in Medicine*, 2004;30:141–151.

[3] Nielsen FÅ. The Brede database: a small database for functional neuroimaging. *NeuroImage*, 2003;19. Presented at the 9th International Conference on Functional Mapping of the Human Brain, June 19–22, 2003, New York, NY. Available on CD-Rom.

[4] Fox PT and Lancaster JL. Neuroscience on the net. *Science*, 1994;266:994–996.

[5] Talairach J and Tournoux P. *Co-planar Stereotaxic Atlas of the Human Brain*. Thieme Medical Publisher Inc, New York, 1988.

[6] Nielsen FÅ and Hansen LK. Modeling of activation data in the BrainMap™ database: Detection of outliers. *Human Brain Mapping*, 2002;15:146–156.

[7] Turkeltaub PE, et al. Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *NeuroImage*, 2002;16:765–780.

[8] Chein JM, et al. Functional heterogeneity within Broca's area during verbal working memory. *Physiology & Behavior*, 2002;77:635–639.

[9] Gitelman DR, et al. Functional anatomy of visual search: Regional segregation within the frontal eye fields and effective connectivity of the superior colliculus. *NeuroImage*, 2002;15:970–982.

[10] Gerlach C, et al. Categorization and category effects in normal object recognition. a PET study. *Neuropsychologia*, 2000;38:1693–1703.

[11] Levy I, et al. Center-periphery organization of human object areas. *Nature Neuroscience*, 2001;4:533–539.

[12] Paulson PE, et al. Gender differences in pain perception and patterns of cerebral activation during noxious heat stimulation in humans. *Pain*, 1998;76:223–229.

[13] Nielsen FÅ and Hansen LK. Experiences with Matlab and VRML in functional neuroimaging visualizations. In Klasky S and Thorpe S, eds., *VDE2000 - Visualization Development Environments, Workshop Proceedings, Princeton, New Jersey, USA, April 27–28, 2000*. Princeton Plasma Physics Laboratory, Princeton, New Jersey, 2000; pages 76–81.