The Brede database: a small database for functional neuroimaging

Finn Årup Nielsen
Informatics and Mathematical Modelling, Technical University of Denmark

Abstract

The "Brede database" provides data for novel information retrieval techniques and automated meta-analyses in functional neuroimaging. The complete database is available on the Web in XML and Matlab format. Results from automated analyses are also available on the Web. Matlab programs are available in the "Brede neuroinformatics toolbox" for analysis and visualization of the data.

Brede neuroinformatics database

Main component: Data from scientific articles reporting Talairach coordinates:
- Cognitive functional neuroimaging experiments (majority)
- Lesion, spatial neglect
- Morphometry, London taxi drivers
- Pathological changes, Alzheimer’s disease

"Poor man’s XML" (pXML, with no attributes and no empty tags): Database kept in a simplified version of XML, distributable on the Internet. Linked to other databases: PubMed, MeSH, SenseLab, fMRIDC.

The database presently consists of data constructed from 85 scientific articles, containing 276 experiments and 1942 locations.

Database components

The database is inspired by the hierarchical structure of BrainMap [1] with scientific articles ("bib" structures) on the highest level containing one or more experiments ("exp" structure), corresponding to a contrast in general linear model analyses, these in turn comprising one or more locations ("loc" structures).

The information on the "bib" level (author, title, ...) is setup automatically from PubMed while the rest of the information is entered manually in a Matlab graphical user interface.

On the "loc" level this includes the 3D stereotactic coordinates in either Talairach or MNI space, the brain area (functional, anatomical or cytoarchitectonic area) and magnitude values such as Z-score and P-value.

Contextual markup

On the "exp" level information such as modality, scanner and behavioral domain are recorded with "external components" (such as "face recognition", "kinetic boundaries", "Alzheimer’s disease" or "BZ site GABA receptor") organized in a directed graph.

Voxel-based analysis

Voxelization: The locations for each "exp" and "bib" structure are voxelized to a volume by convolving each location with a Gaussian kernel [8, 11].

The voxelization equation for a specific experiment \( e \) from its set of locations \( \{ x_i : i \in \mathcal{E}_e \} \) is

\[
p(x|e) = \sum_{i \in \mathcal{E}_e} (2\pi \sigma^2)^{3/2} \exp \left[ -\frac{(x - x_i)^2}{2\sigma^2} \right] P(I),
\]

where there is an equal weighting over locations \( P(I) = 1/|\mathcal{E}_e| \) and the kernel width is fixed at \( \sigma = 1 \) cm. Alternatively, it can be optimized with leave-one-out cross validation [8]. The voxelization for a specific "bib" item \( b \) is constructed as the average over its "exp"

\[
p(x|b) = \sum_{e \in \mathcal{E}_b} p(x|e) P(e),
\]

with priors taken as uniform over experiments \( P(e) = 1/|\mathcal{E}_e| \).

Voxelization can be regarded as the inverse operation of finding the maxima in an image:

Local Maximum / Center of gravity

Volume

Locations
Static web pages are generated from the "exp" and "bib" structures with Corner Cube visualization [10] as PNG and VRML files and hyperlinks to PubMed and fMRIDC [12]. The combined set of volumes are converted to matrices and the following analyses are performed automatically:

- Multivariate analyses are performed such as singular value decomposition (SVD), independent component analysis (ICA) and non-negative matrix factorization (NMF) [4].

- Finding related volumes: Sorted lists with related volumes are found for each individual volume as well as with respect to the SVD eigenimages and the results of the ICA and NMF [9].

- Asymmetry: Counting the number of locations in left and right hemisphere and comparing the counts to a binomial distribution gives a laterality index.

- Novelty: Comparing each volume to mean or nearest volume gives an estimate of novelty or "outliersness": How different a volume is from the rest.

Ad hoc search can obtain the closest locations to a user-specified coordinate or the closest experiments to a user-specified set of locations.

Items in the database are identified with unique numbers and the type of identifier is given a unique string, e.g., "WOBIB: 27" for an Epstein and Kanwisher paper, thus allowing Internet search engines to identify the phrase.

A webpage constructed from the "exp" entry (WOEXP: 196) of an experiment reported in [8] - a morphometry study of London taxi drivers: Listed with Talairach coordinates in the hippocampus and a Corner Cube visualization where two locations are shown as glyphs together with an iso surface as wireframe in the volume from voxelization.

The first part of the list with related volumes are shown at the bottom.

The "face recognition" external component (WOEXT: 23) with links to experiments marked up as "face recognition", here presently experiments from [5, 3, 2].

Database distributed as XML and Matlab files. Most recent update available from http://hendrix.imm.dtu.dk/services/jernej/brede/

Matlab programs for manipulating, analyzing and visualizing the data are available in the "Breden Bioinformatics toolbox" [7]. http://hendrix.imm.dtu.dk/software/brede/

Lars Kai Hansen, Finn Anup Nielsen is funded by the European Union project MAPAWAMO.