Modeling of locations in the BrainMap database: Detection of outliers

F. Å. Nielsen, L. K. Hansen and U. Kjems

Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby, Denmark

Introduction

We make statistical meta-analytic modeling of locations from BrainMap [1] describing the relationship between the 3D Talairach coordinates and the anatomical labels, which enable us to make automated novelty detection [2] and spot outliers, e.g., infrequent or non-conforming terminology or transcription errors.

Method

We downloaded most of the webpages from the BrainMap and extracted the 3D Talairach coordinates $\mathbf{x}$ as well as the anatomical label (BrainMap: "Lobar Anatomy"). Each anatomical label was split into all subset of phrases, e.g., the string "midline occipital lobe" generated an event in the classes of "midline", "occipital", "lobe", "midline occipital", "occipital lobe" and "midline occipital lobe". We next constructed a probability density model in the 3D Talairach space $\mathbf{x}$ by conditioning on the word/phrase w class: $p(\mathbf{x}|w)$. For probability density modeling we used a variation of a Specht kernel estimator [3], where the width of the Gaussian kernel (σ^2) is optimized by leave-one-out (LOO) cross-validation. Our implementation is based on a fast one-dimensional Newton optimization of the leave-one-out cost function (negative log probability),

$$E(\sigma^2, w) = -\sum_{n=1}^{N_w} \log p_{-n}(\mathbf{x}_n|\sigma^2, w), \tag{1}$$

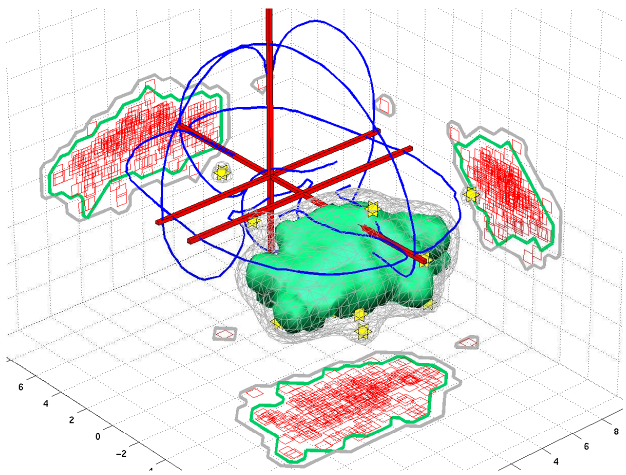
where $\mathbf{x}_n$ is the 3D Talairach coordinate of the n'th location with label w, N_w is the number of locations labeled w, and the density based on all examples except the n'th is given by

$$p_{-n}(\mathbf{x}|\sigma^2, w) = \frac{1}{N_w - 1} \sum_{n' \neq n}^{N_w} \left(2\pi\sigma^2\right)^{-3/2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{x} - \mathbf{x}_{n'})^2\right). \tag{2}$$

As the probability density estimation will be affected by outliers we used a two-stage ad hoc strategy: In the first stage we use all coordinates within the class to obtain the probability density. In the second stage we exclude the 5% most unlikely coordinates and estimate the probability density on the remaining 95%. Novelty detection is implemented by ranking the estimated density value $p(\mathbf{x})$ of the locations. We manually investigated the top-ranked outliers.

Result

Modeling 3935 location with 1231 word/phrases and investigating the 21 highest ranked outliers we found some were due to entry errors, e.g., millimeters/centimeters interchanged and some were transcription errors on the level of the article. Other outliers was due to infrequent terminology, e.g. the word "lobe" is typically used in connection "parietal" and "occipetal"; seldomly in a phrase with "temporal", "limbic" or "subgyral". The figure presents an example with the density formed by modeling 294 locations labelled "cerebellum". The first level density is shown as wireframe, and the second level trimmed density is shown as a rendered surface. Note that two isolated "blobs" that were created by isolated outlying locations were eliminated going from the first to the second level density, — one of them representing a location that were among the 21 investigated outliers.



Conclusion
We find the Specht kernel estimator simple and efficient to flexibly describe a set of BrainMap locations and performing well in detecting outliers. This will make it a good candidate for modeling relationship between location and behaviour.

Acknowledgement
BrainMap, Danish Research Councils and Human Brain Project P20 MH57180, E. Tulving, M. Corbetta, D. J. Brooks, H. Jenkins and B. Shaywitz and Daniela Balslev.

References
1. Fox, P. T., Lancaster, J. L., Science, November 1994, 266(5187):994–996.
2. Bishop, C. M., IEE Proceedings — Vision, Image and Signal Processing, August 1994, 141(4):217–222.
3. Specht, D. F., Neural Networks, 1990, 3(1):109–118.