Published in final edited form as:

Cortex. 2018 November ; 108: 276–278. doi:10.1016/j.cortex.2018.09.004.

Estimating the Statistical Significance of Spatial Maps for Multivariate Lesion-Symptom Analysis

Grigori Yourganov¹, Julius Fridriksson², Chris Rorden¹

¹Department of Psychology, University of South Carolina, Columbia, SC, USA

²Center for the Study of Aphasia Recovery, Department of Communication Sciences and Disorders, University of South Carolina, Columbia, SC, USA

Multivariate analysis methods are becoming increasingly popular to study the relationship between behavior and neuroimaging data. In recent years, multivariate methods have been developed specifically for predicting stroke-related impairments using brain imaging data (Hope et al., 2013; Smith et al., 2013; Zhang et al., 2014; Yourganov et al., 2016; Pustina et al., 2017). An important theoretical advantage of multivariate methods over univariate methods (such as voxel-based lesion-symptom mapping, or VLSM; Bates et al., 2003) is the fact that behavioral functions often recruit spatially distributed cortical networks; patients with damage to distinct parts of these networks will serve as counter-examples for VLSM (Karnath & Smith, 2014). On the other hand, multivariate training can identify these spatially distributed patterns of injury to provide more accurate out-of-sample predictions (see e.g., Schmah et al., 2010; Yourganov et al., 2014).

However, an advantage of mass-univariate methods is the ease of interpretation of the statistical parametric maps generated by the analysis. For example, in VLSM a statistical test (such as t test) is performed at every voxel, to see if there is a significant difference in behavior between the participants with and without damage in that voxel. As a result, we get a spatial map with voxelwise t values (and corresponding p values). Current voxelwise multivariate methods (such as SVR-LSM; Zhang et al., 2014) lack this feature. A linear multivariate model does indeed assign a weight to each voxel, which reflects the importance of this voxel in predicting behavior. However, this importance is relative to other voxels, and the weights are essentially scale-free, with unclear statistical significance.

To threshold the maps, one can use the permutation technique (Nichols & Holmes, 2002), where the behavioral scores are randomly permuted thousands of times to create the distribution of the test statistic under the null hypothesis and use it to derive the voxelwise p -values. However, this approach is very computationally intensive, especially when combined with cross-validation. For example, our predictive framework (described in Yourganov et al., 2017) utilizes a leave-one-out scheme, where the linear support vector regression (SVR) model is trained using the data from all but one participant, and this model is used to predict the behavioral score of the left-out participant. After we iterate through all participants, the

correlation between actual and predicted behavioral scores is a measure of generalizability of our analysis to out-of-sample data. Using a contemporary Mac Book Pro i7 laptop, it takes approximately 80 minutes to run it on our sample of lesion maps obtained for 156 participants (each map consisting of 437,123 voxels that were damaged in at least 10 participants). Using this setup, running 1000 permutations would take 55 days.

Fortunately, significance of voxelwise predictive weights can be estimated within the leave-one-out scheme at no additional computational cost, using the theory developed for jackknife resampling (Efron & Tibshirani, 1994, p. 141–151). Leave-one-out cross-validation and jackknife resampling have been developed for different purposes, but the underlying procedure is identical: iteratively, one observation is left out of the sample and the test statistic is computed on the remaining observations. The standard error of the test statistic is estimated as

$$se_{jack} = \sqrt{\frac{N-1}{N} \sum_{i=1}^N (\theta_i - \theta_{av})^2},$$

where N is the number of observations, θ_i is the estimate of the test statistic on the sample with the i th observation left out, and θ_{av} is the average of such estimates. This expression is similar to the standard deviation of θ_i computed from N independent samples, but it is scaled by the factor of $\sqrt{N-1}$ to account for the fact that our jackknife samples are greatly overlapping rather than independent.

We propose converting the predictive weights into voxelwise t scores using the following procedure. Each leave-one-out iteration gives us a set of voxelwise predictive weights. For each given voxel, we compute the mean predictive weight across the iterations, as well as the jackknife estimate of standard error; the ratio of the mean to the standard error is the one-sample t statistic. The corresponding voxelwise p -values can be corrected for multiple comparison using Bonferroni or False Discovery Rate (FDR) thresholding (or permutation-based thresholding, if one is willing to invest in computing a large number of permutations). This thresholding could be used as feature selection (to be evaluated on an independent test set), keeping in mind the possibility that no voxels survive (Pereira et al., 2009), indicating the situation when multivariate model could yield good prediction, but each variable, taken by itself, is not significantly contributing to it.

Figure 1A shows the distribution of our voxelwise t values when we estimate Western Aphasia Battery naming scores (Kertesz, 1982) for our sample of 156 chronic-stroke participants using the linear SVR algorithm (Pearson's correlation coefficient between actual and predicted naming scores is 0.659, $p < 10^{-20}$). In comparison, the dashed line shows the distribution of t values obtained with VLSM. The voxelwise t maps for linear SVR and VLSM are shown on Figure 1B. The lesion size was corrected by using direct total lesion volume control (Zhang et al., 2014) in SVR analysis, and by regressing it out of the naming scores in VLSM analysis. It should be noted that when the model was retrained using only the voxels that were significant on $p < 0.05$ level (about 30% of voxels survived the FDR correction), there was a slight improvement of prediction from 0.659 to 0.674.

We have demonstrated that leave-one-out framework serves a dual purpose: out-of-sample prediction, and estimation of statistical significance of each voxel's contribution to the multivariate predictive model. The statistical parametric maps obtained with the jackknife resampling can be used to localize the distributed cortical networks where post-stroke damage is associated with behavioral impairment. Our procedure is inspired by the NPAIRS framework (Strother et al., 2002) which uses split-half resampling to analyze functional MRI data. The MATLAB code for our procedure is included into the NiiStat software package, a versatile tool for analyzing the brain-behavior relationships using multiple types of neuroimaging data (available online at <https://github.com/neurolabusc/NiiStat>).

ACKNOWLEDGMENTS

This research was funded by National Institute on Deafness and Other Communication Disorders Grants DC008355 and DC009571 and grant NIH NIDCD DC014664.

REFERENCES

- Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, & Dronkers NF (2003). Voxel-based lesion-symptom mapping. *Nature neuroscience*, 6(5), 448. [PubMed: 12704393]
- Efron B, & Tibshirani RJ (1994). *An introduction to the bootstrap*. CRC press.
- Hope TM, Seghier ML, Leff AP, & Price CJ (2013). Predicting outcome and recovery after stroke with lesions extracted from MRI images. *NeuroImage: Clinical*, 2, 424–433.
- Karnath HO, & Smith DV (2014). The next step in modern brain lesion analysis: multivariate pattern analysis. *Brain*, 137(9), 2405–2407. [PubMed: 25125587]
- Kertesz A (1982). *Western aphasia battery test manual* (Psychological Corp, New York)
- Nichols TE, & Holmes AP (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*, 15(1), 1–25. [PubMed: 11747097]
- Pereira F, Mitchell T, & Botvinick M (2009). Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage*, 45(1), S199–S209. [PubMed: 19070668]
- Pustina D, Avants B, Faseyitan OK, Medaglia JD, & Coslett HB (2017). Improved accuracy of lesion to symptom mapping with multivariate sparse canonical correlations. *Neuropsychologia*.
- Schmah T, Yourganov G, Zemel RS, Hinton GE, Small SL, & Strother SC (2010). Comparing classification methods for longitudinal fMRI studies. *Neural computation*, 22(11), 2729–2762. [PubMed: 20804386]
- Smith DV, Clithero JA, Rorden C, & Karnath HO (2013). Decoding the anatomical network of spatial attention. *Proceedings of the National Academy of Sciences*, 110(4), 1518–1523.
- Strother SC, Anderson J, Hansen LK, Kjems U, Kustra R, Sidtis J, Frutiger S, Muley S, LaConte S, & Rottenberg D (2002). The quantitative evaluation of functional neuroimaging experiments: the NPAIRS data analysis framework. *NeuroImage*, 15(4), 747–771. [PubMed: 11906218]
- Yourganov G, Schmah T, Churchill NW, Berman MG, Grady CL, & Strother SC (2014). Pattern classification of fMRI data: applications for analysis of spatially distributed cortical networks. *Neuroimage*, 96, 117–132. [PubMed: 24705202]
- Yourganov G, Fridriksson J, Rorden C, Gleichgerrcht E, & Bonilha L (2016). Multivariate connectome-based symptom mapping in post-stroke patients: networks supporting language and speech. *Journal of Neuroscience*, 36(25), 6668–6679. [PubMed: 27335399]
- Zhang Y, Kimberg DY, Coslett H, Schwartz MF, & Wang Z (2014). Multivariate lesion-symptom mapping using support vector regression. *Human brain mapping*, 35(12), 5861–5876 [PubMed: 25044213]

Figure 1. Statistical maps obtained from analyzing naming scores with SVR and VLSM. A: Distribution of the t scores in the statistical maps (SVR: solid; VLSM: dashed). B: The t maps for SVR (top) and VLSM (bottom).