Considerations When Using Single-Trial Parameter Estimates in Representational Similarity Analyses

Jeanette A. Mumford, PhD

Department of Psychology
University of Texas at Austin
Austin, Texas

Introduction

Traditional analyses of functional magnetic resonance imaging (fMRI) data compare mean blood-oxygen-level–dependent (BOLD) activations during different tasks to determine where brain activation differences occur. More recently, the patterns of activation within regions of the brain have been studied with keen interest. In them, the input patterns consist of either the test statistics or the activation estimate magnitudes associated with a task over space (Haxby, 2001; Carlson et al., 2003; Pereira et al., 2009; Kriegeskorte et al., 2008; Kriegeskorte, 2011). The two most common pattern analyses are multivariate pattern classification analyses and pattern similarity analyses, which are commonly referred to as representational similarity analyses (RSAs).

Multivariate pattern classification analyses use activation patterns, either over the entire brain or within a region of interest (ROI). They use a cross-validation approach, where part of the data is used to train a model that discerns between different task patterns; then, this fitted model is used to predict the task associated with another set of independent patterns, ignoring their labels. The performance of the model is assessed by the accuracy of the predictions. The models used include regularized logistic regression and support vector machine-based classification algorithms (Carlson et al., 2003; Pereira et al., 2009). RSAs correlate activation patterns over space with different tasks to assess the similarity in activation patterns. One can then compare the RSA values, which consist of all possible pairwise correlations between tasks, across different regions of interest and between different imaging modalities (Kriegeskorte et al., 2008; Kriegeskorte, 2011). For example, in Kriegeskorte et al. (2008), RSA-based networks were used to compare human and primate network similarities. The activation patterns used in classification and similarity analyses can be based on estimates for individual trials, as well as windows of time surrounding a trial, and be aggregated over many trials (Xue et al., 2010; Mumford et al., 2012; Turner et al., 2012).

Because RSA analyses are a newer type in the field, the methods for data analysis have not yet been perfected. The focus here is on RSA analyses that use single-trial activation patterns to generate similarity matrices. For example, a single run of the paradigm might include presenting images of 30 faces and 30 houses, and the goal is then to compute a 60 × 60 similarity matrix to be used in the analysis. This type of approach has been used in the past for designs in which the trials are temporally spaced (Kriegeskorte et al., 2008; Jenkins and Ranganath, 2010; Xue et al., 2010; Ritchey et al., 2012). Recent work has developed methods for obtaining single-trial parameter estimates for faster event–related designs that have improved performance in classification based analyses (Mumford et al., 2012). Therefore, it is only natural that these models be used in RSA-based analyses. This chapter discusses the methods for obtaining single-trial parameter estimates as well as possible problems that can arise when these estimates are used in RSA analyses.

Modeling Single-Trial Activation Patterns

Obtaining single-trial parameter estimates can be difficult in studies where the stimuli are presented with a short interstimulus interval (ISI). This is because the blurring and delayed nature of the BOLD fMRI signal makes it difficult to isolate a signal that is unique to a specific trial. Previous work introduced a new approach for estimating single-trial parameter estimates in fast event–related designs: using a separate model for each trial and iteratively estimating the pattern for each trial separately. This method is referred to as least squares single (LSS) and is illustrated in the right panel of Figure 1 (Mumford et al., 2012). This method was shown to produce less

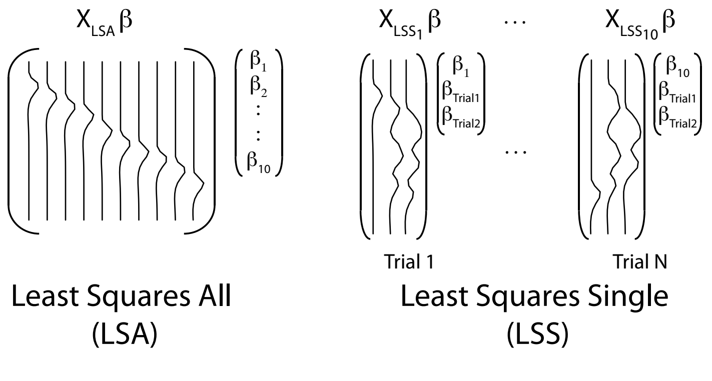
$X_{LSA}\beta$
β_1 β_2 : : β_{10}
$X_{LSS_1}\beta$
β_1 β_{Trial1} β_{Trial2}
…
$X_{LSS_{10}}\beta$
β_{10} β_{Trial1} β_{Trial2}
Trial 1
Trial N
Least Squares All (LSA)
Least Squares Single (LSS)

Figure 1. Model illustration for LSA and LSS. In both cases, trial-specific activations are estimated for each of 10 trials, and the model is run in a voxelwise fashion. The left panel shows LSA, which estimates all trials simultaneously in a single regression, and the estimates β_1, …, β_{10}, which represent the activation magnitudes for each trial. The right panel shows LSS, where each trial's activation is estimated in a separate model. The first regressor represents the trial of interest, and the two additional regressors model the remaining trials according to trial type (in this case, two). Only the estimates for the first parameter are retained from each model.

NOTES

variable estimates than the more standard approach, which estimates all trials simultaneously in a single model. The standard approach is referred to as least squares all (LSA) and is illustrated in the left panel of Figure 1. Although the behavior and performance of LSS and LSA estimates were studied within the context of pattern classification analyses, they have not been studied within the RSA setting.

Factors That Influence RSA Measures

Noise differences

Since the RSA measure is based on correlations between activation patterns, it is susceptible to noise differences. For example, if similarity matrices are obtained for multiple ROIs, and one is interested in studying the correlation of the RSA matrices between ROIs (Kriegeskorte et al., 2008; Diedrichsen et al., 2011), then noise variations between regions can induce false-positive differences. For example, if one region has noisier data than another, but the theoretical correlations of the within-region RSA are identical for the two regions, then the noisier region will have smaller RSA values than the other. This problem was discussed by Kriegeskorte et al. (2008), who concluded that a rank-based correlation is necessary to avoid correlation magnitude differences when comparing RSAs between regions. It was also addressed by Diedrichsen et al. (2011), who used a random-effects model to estimate the RSA correlations within regions, adjusting for noise differences. Theoretically, this method would allow for directly comparing RSA values between regions, because it would remove any biases induced by noise differences.

Variability differences

Although spatial noise differences and their impact on RSA estimates have been studied, variability differences across trial pattern estimates also can induce false-positive RSA differences within a region of interest. For example, assume that a study has 30 presentations of each of two trials (type 1 and type 2). The T1 trials have a longer duration than the T2 trials, and the RSA study of interest is how the distributions of within-trial RSA values compare between these two trials. It is well known that longer trials yield more efficient estimates of BOLD activation when using the general linear model (GLM), so trial A will have less variable activation-pattern estimates than trial B (Smith et al., 2007). Even if the true within-trial correlation were identical for both trial types, the higher variability in T2's estimates would lead to within-trial RSA values lower for T2 than for T1. This is not the only case where variability differences may occur between trials, within a region of interest. If the ISI (duration of the baseline task surrounding each trial) differs between trials, shorter trials will also have noisier activation-pattern estimates.

Model influence

A less studied phenomenon is how the model itself can determine correlations between activation patterns. This is the case for both LSS and LSA methods for obtaining activation patterns for single trials. Figure 2 illustrates the true RSA values between 60 trials (the first 30 trials are type 1 and second set of 30 are type 2), where a correlation of 0 is assumed between all trials, and the ISI is equal for all trials and 2 s on average (left panel). The middle and right panels show the theoretical correlations

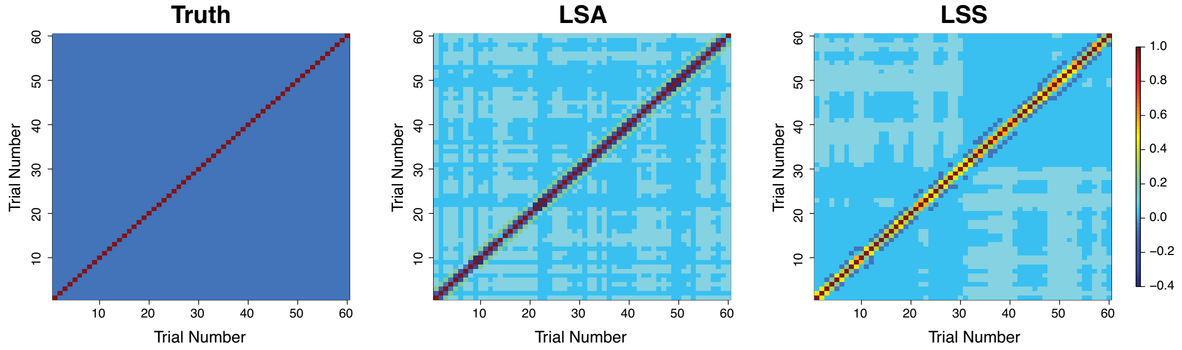

Figure 2. RSA estimate comparisons for the null case. The run consisted of 20 trials, each 2 s long with a random ISI between 2 s and 4 s. There were two trial types: 30 of trial type 1, followed by 30 of trial type B. Trials are ordered in the matrices as they were presented in time, starting from trial 1 (bottom left corner) to trial 60 (top right corner). Each cell in the matrix corresponds to the correlation between activation patterns between two trials. The left panel illustrates the true RSA, which in this case of the null is set to an identity matrix (1s on diagonal and 0s elsewhere). The middle and left panels are the LSA- and LSS-based RSA matrices that were theoretically derived.

between the activation patterns when using LSA and LSS, respectively. In these graphs, the trials are ordered temporally from first to last, from the bottom left corner to the top right corner of the plot. The first off-diagonal of the plot, therefore, indicates the lag 1 correlations between trials.

As this figure shows, for both LSS and LSA, there is a bias in the lag 1 correlation. In LSS, trials that are temporally adjacent (lag 1) are anticorrelated, since the collinearity in the model pushes the estimate of one trial in the opposite direction of its collinear neighbors. In contrast, LSA has a positive bias that occurs because the trials are modeled separately and, hence, are not adjusted for each other statistically and thus describe similar parts of the data. LSS shows a blocked pattern caused by a collinearity effect in the LSS model that occurs when the neighbors of a trial match each other, but not necessarily the trial in the middle. This causes a weak negative correlation between that trial and all trials of the neighbors' type, owing to a collinearity between the single trial and the nuisance regressor for the trial of the neighbors' type. The fact that trials are blocked in Figure 2 causes a weak negative correlation between all trials of the same type, though correlations between trial types are comparatively more positive—a counterintuitive result. These patterns are described in more detail for a variety of trial orderings and ISI durations below.

Trial order and ISI duration

A series of simulations were used to study how the patterns in Figure 2 vary as a function of the ordering of the trials and the ISI duration. Trials were either blocked (all trial 1, followed by all trial 2), alternated (T1, T2, T1, etc.), or randomly generated for each run. Situations where the ISIs matched between trials and when the ISI was longer for T2 were investigated. The ISIs were randomly drawn from continuous uniform distributions, where $U(a,b)$ represented random draws between a and b. The TR used in the simulations was 2 s.

To start, the behavior of the correlation as a function of lag was estimated. Figure 3 shows the mean correlation as a function of lag for ISIs ranging between 2 s and 7 s, where ISIs matched for the two trial types. The left panel shows the patterns for LSA, in which for an ISI of 2 (red line), the correlation alternates between negative and positive. This pattern results from collinearity pushing neighboring trials in opposite directions. Thus, at a lag of 1, the correlation would be negative, but at a lag of 2, there is a positive correlation (since both trials would be pushed in the same direction by their common collinear neighbor). This effect diminishes as the ISI increases, since the collinearity problem is reduced. Even so, the mean correlation is never the true correlation of 0 but hovers around 0.1—the result of the influence of the model itself.

The LSS model shows positive correlations for early lags owing to the overlap in the trials, which are not statistically adjusted for each other and hence describe the same variability in the data. The dip around a lag of 3 when the ISI equals 2 results from the peak of one trial meeting the poststimulus undershoot of another, a difference of approximately 6 s.

The results in Figure 4 compare within-trial correlations (red and green) to between-trial correlations (blue) when the ISIs match and when T2's ISIs are longer

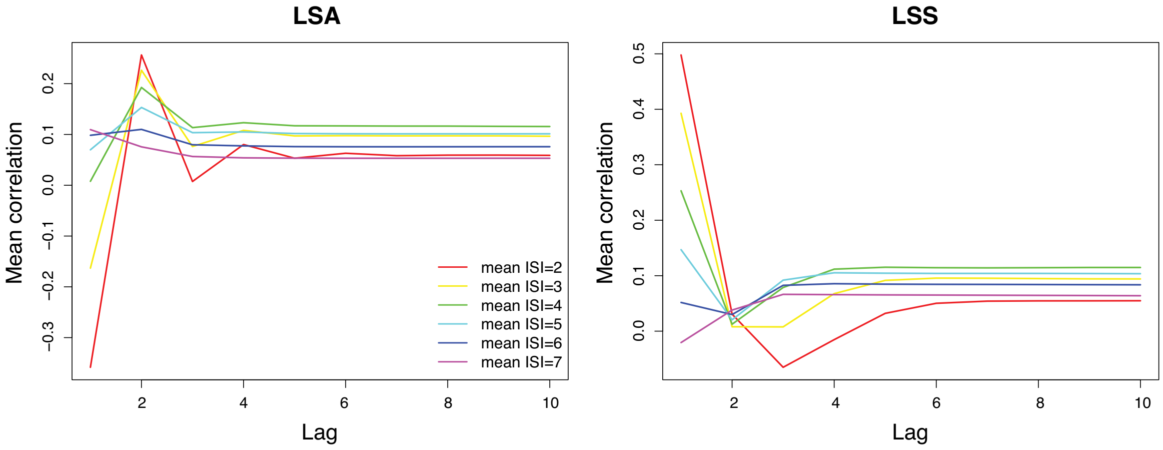

Figure 3. Averaged correlations as a function of lag for LSA (left) and LSS (right) for varying ISIs when trials are blocked. LSA tends to alternate between positive and negative due to collinearity; in contrast, LSS tends to have high positive correlations at early lags, followed by a negative correlation when the peak of one trial meets with the poststimulus undershoot of another trial.

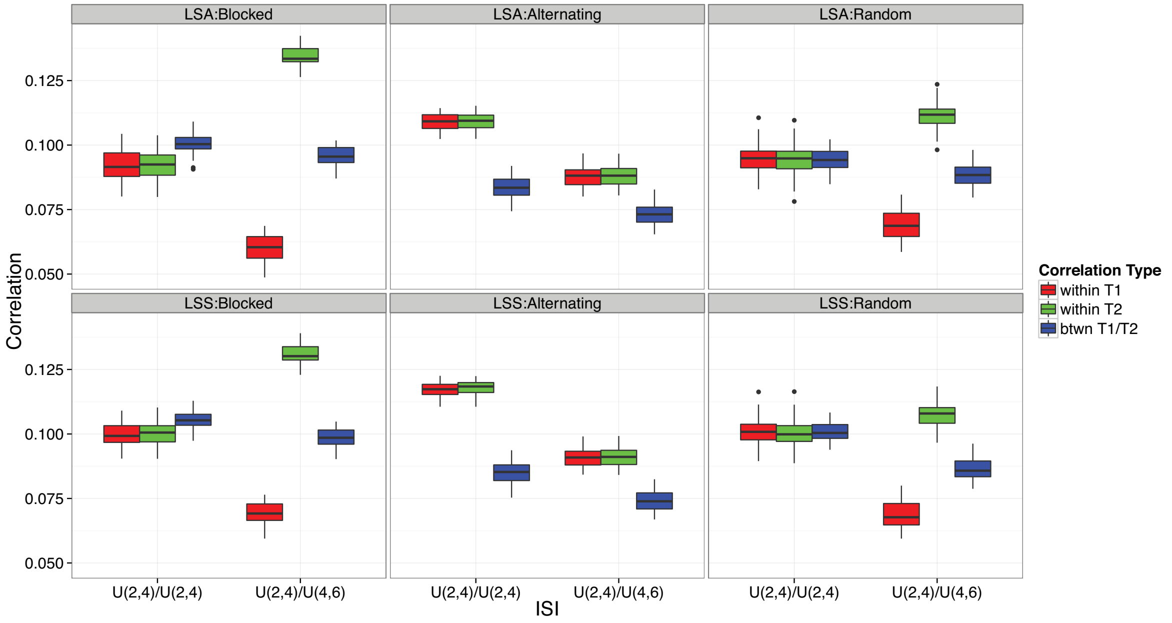

Figure 4. Distributions of within-trial and between-trial correlations for 100 simulated data sets for varying ISI settings and trial orderings for LSA (top) and LSS (bottom). The setting of equal ISIs with random trial orderings appears to remove biases, assuming that a different randomization is used for each subject.

for blocked (left), alternating (middle), and random (right) trial orderings. Significantly, in the random setting (right), each run had a different randomization. The top row shows the LSA results, and the bottom shows LSS ones. The alternating blocked designs illustrate that, even with the ISIs matching, the within-trial correlations are not comparable with the between-trial correlations. In the alternating LSA, the within-trial correlations are larger; in the blocked case, they are smaller, as noted in Figure 1. When the ISIs differ between trials, there is not much impact in the alternating case since, on average, each trial has the same amount of fixation or rest on either side of it. In both the blocked and random cases, the second trial type will always have more fixation surrounding it, which decreases the collinearity problem for trials of type 2, removing some bias in the correlation estimate. The only case where the correlations are comparable for both LSA and LSS is when the ISIs match and the trial order is randomized.

Randomly ordering trials and matching the ISI seem to be key for both LSS and LSA; however, it is important that the trials be randomly ordered for each run of the analysis. In reality, it is common to generate one or two trial orderings found to be efficient for the design and to use these for all subjects. But when this is done, slight biases in ISI may occur. In 100 simulated data sets with 40 subjects in each study, more than half of the data sets exhibited a statistically significant difference between at least one of the pairwise comparisons (within T1 versus within T2; within T1 versus between T1/T2; or within T2 versus between T1/T2). Figure 5 illustrates the distribution of RSA values for a single one of these simulated data sets. It shows that the within-T2 correlations are significantly larger than both within-T1 and between-T1/T2 ones. Although the magnitude of the difference is quite small, it is common to look only at the *p* values of these tests, so this would not be taken into consideration.

Possible Remedies and Considerations

One possible fix for this problem is to compute RSA matrices by using comparisons of patterns estimated with different models. For example, if there are two runs of a study for a single subject, correlate only a single trial's pattern from run 1 with trials in run 2. Since the biases discussed above were all imposed by the models used, this will prevent that bias from entering the analysis. Notably, however, this will not solve problems where the ISI was not equal across all trial types.

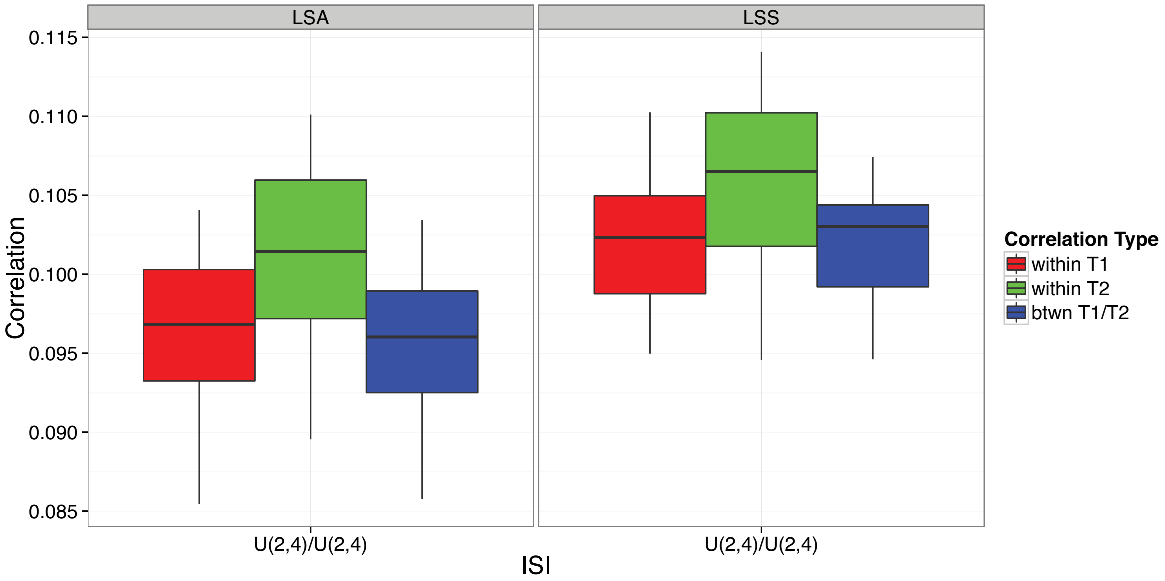

Figure 5. When the same randomization is used across all subjects, unusual patterns can emerge that cause small yet statistically significant differences in RSA distributions. In this case, the within-T2 similarities appear larger than within-T1 and between-T1/T2 similarities, when theoretically, they all equal 0.

Before running a study, it is fairly easy to take the proposed paradigm and perform simulations like the ones run here to investigate possible biases that may enter the analysis. However, using between-run correlations should remedy many of the problems. What is important is to recognize that paradigms that work very well when studying activation magnitudes and their differences may not work as well in the RSA setting. Although only ISI differences were studied here, the same problems would arise if there were differences in the duration of the stimuli. Where stimulus presentation is controlled by the reaction time of the subject, further analysis strategies would need to be developed to control for those differences.

References

Carlson TA, Schrater P, He S (2003) Patterns of activity in the categorical representations of objects. J Cogn Neurosci 15:704–717.

Diedrichsen J, Ridgway GR, Friston KJ, Wiestler T (2011) Comparing the similarity and spatial structure of neural representations: a pattern-component model. Neuroimage 55:1665–1678.

Haxby JV (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425–2430.

Jenkins LJ, Ranganath C (2010) Prefrontal and medial temporal lobe activity at encoding predicts temporal context memory. J Neurosci 30:15558–15565.

Kriegeskorte N (2011) Pattern-information analysis: From stimulus decoding to computational-model testing. Neuroimage 56:411–421. Available at http://linkinghub.elsevier.com/retrieve/pii/S1053811911000978.

Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis—connecting the branches of systems neuroscience. Front Sys Neurosci 2:4.

Mumford JA, Turner BO, Ashby FG, Poldrack RA (2012) Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage 59:2636–2643.

Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 45:S199–S209.

Ritchey M, Wing EA, LaBar KS, Cabeza R (2012) Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. cerebral cortex. 2012 Sep 11. [Epub ahead of print]

NOTES

Smith S, Jenkinson M, Beckmann C, Miller K, Woolrich M (2007) Meaningful design and contrast estimability in FMRI. Neuroimage 34:127–136.

Turner BO, Mumford JA, Poldrack RA, Ashby FG (2012) Spatiotemporal activity estimation for multivoxel pattern analysis with rapid event-related designs. Neuroimage 62:1429–1438.

Xue G, Dong Q, Chen C, Lu Z, Mumford JA, Poldrack RA (2010) Greater neural pattern similarity across repetitions is associated with better memory. Science 330:97–101.