ROC methods in fMRI with real data using repeated trials: Limitations and Improvements

R. R. Nandy[1], D. Cordes[1]

[1]Department of Radiology, University of Washington, Seattle, Wa, United States

Introduction

The most popular and useful tools to assess the efficiency of different fMRI post-processing algorithms are methods based on receiver operating characteristic (ROC) [1]. The method is based on simulation and is limited in scope primarily due to difficulties in incorporating the spatial dependence among neighboring voxels in the simulation. A modified method has been suggested previously using real fMRI data by Le & Hu [2]. We propose an improvement of Le & Hu's method using some mathematical relationships and also demonstrate how to reconstruct the conventional ROC curves from the modified curves.

Theory and methods

The ROC curve is a plot of True Positive Fraction (TPF) against False Positive Fraction (FPF) for detected voxels in an fMRI post processing algorithm. To locate the truly active voxels for the task of interest from real fMRI data, Le and Hu proposed a data driven method. In this method, the subject is presented with a standard periodic ON/OFF paradigm repeated for several periods. True positives are then estimated using a t-test with a very high degree of confidence (p-value of the order of 10^{-6} or less). The voxels detected to be active at this threshold are postulated to be the true positives. A second set of data is collected with a similar paradigm and TPF is estimated using the voxels postulated to be true positives from the first dataset. ROC curves are then plotted for the post-processing algorithm of interest. Since a very high level of confidence is used to estimate the locations of the active voxels, for the sake of simplicity, we can assume that none of the voxels are mistakenly detected to be active. However, the price to pay to achieve such a high degree of confidence is missing out truly active voxels, since, for a fixed sample, the probability of Type II error increases with the reduction of the probability of Type I error. Mathematically this can be formulated as follows. Let N be the total number of voxels among which N_T voxels are true positives and N_F voxels are inactive. The N_T true positive voxels can be further split as N_{T1} voxels that are detected to be active from the first dataset and N_{T2} voxels that are not detected and incorrectly declared to be inactive. Hence, the voxels can be classified broadly into three groups; the first being the collection of voxels that are truly active and labeled as active (T1), the second being the collection of voxels that are truly active and labeled as inactive (T2) and finally the collection of voxels that are truly inactive and labeled as inactive as well (F). By construction, the group T1 consists of voxels with highest level of activity and is completely known. The groups T2 and F, however, are not known individually but the collection of voxels combining these two groups is known. To plot the ROC curves using Le & Hu's method, it is necessary to incorrectly assume that there are N_{T1} true positive voxels and ($N_F + N_{T2}$) inactive voxels. Let $\alpha_1 = N_{T1}/N_T$, $\alpha_2 = N_{T2}/N_T$, $\beta_1 = N_{T2}/(N_{T2}+N_F)$ and $\beta_1 = N_F/(N_{T2}+N_F)$. For any threshold u, $p_1(u)$, $p_2(u)$ and $p_3(u)$ are, respectively, defined as the probabilities of detecting a voxel as active at threshold u for a test statistic from the three groups of voxels classified above. The true ROC curve is a plot of $h(u)=\alpha_1 p_1(u)+\alpha_2 p_2(u)$ against $g(u)=p_3(u)$ for all possible thresholds, whereas Le & Hu's ROC curve is a plot of $h^*(u)=p_1(u)$ against $g^*(u)=\beta_1 p_1(u)+\beta_2 p_3(u)$. The ratio of the slopes of the true ROC curve and Le & Hu's curve is given by $R = [h'(u)/\, g'(u)]/\, [h^{*'}(u)/\, g^{*'}(u)]$. If we assume that $p_1(u)$, $p_2(u)$ and $p_3(u)$ all belong to the same Gaussian distribution except for a difference of means, under certain regularity conditions (a detailed description is beyond the scope of this abstract), it can be mathematically shown that $R<1$ near the origin. This ensures that near the origin (which is the most critical part of the ROC curve in fMRI), Le & Hu's ROC curve lies above the true ROC curve until they meet. Thus a conservative choice of the initial threshold as proposed by Le & Hu may not be appropriate. Instead, we suggest using a threshold such that the proportion of voxels detected to be active in the first dataset is equal to the expected proportion of truly active voxels. Such a scheme is advantageous for several reasons. To begin with, in most cases, even though we do not know the exact locations of the active voxels *a priori*, we may have a good idea about the expected number of active voxels and this is all we need to determine the threshold. It is also possible to estimate the number of active voxels using p-values [3]. When such a threshold is applied to the first dataset, some of the truly active voxels will not be detected and an equal number of inactive voxels will be incorrectly labeled as active. While processing the second dataset, if the post-processing method being evaluated by ROC method tends to pick up the inactive voxels labeled as active before the active voxels labeled as inactive, the ROC curve with the modified threshold will be pushed upward. On the other hand, if it is the opposite, the curve will be pushed downward. This means that with our choice of threshold there is no obvious bias in the modified ROC curve. It should also be mentioned that the true ROC curve can be reconstructed from Le and Hu's curve using a method based on p-values which has been described in a recent article [3].

Results

To compare the true ROC curve with Le & Hu's ROC curve, we have used simulated data. To construct the simulated data, pure resting state data is used as noise which is added to the simulated activation. We have used two sets of resting-state data for 10000 voxels and a gamma-convolved boxcar function (four periods, each period being 60 timeframes long) has been added to the same 10000 voxels for both the datasets. The amplitude of the boxcar function has been adjusted so that the CNR is 0.6. The post-processing method for the second dataset to generate the ROC curve has been the standard univariate regression analysis with Fourier basis. A p-value threshold of 10^{-6} is used with the first dataset to identify the active voxels according to Le & Hu's method. In Figure 1, we plot Le & Hu's curve, the true ROC curve as well as the curve proposed by us using a modified threshold. In Figure 2, we plot the estimated ROC curve using our reconstruction method with the same data. For real data, we have used a standard on-off motor paradigm (six periods, each period being 40 sec. long) which has been repeated in identical environment. In Figure 3, ROC curves for a standard univariate regression analysis (with Fourier basis functions of same periodicity as regressors) are plotted from the second dataset for different choices of thresholds for the t-test applied to the first dataset. We have labeled the proportion of voxels initially selected to be true positives for the corresponding threshold as "p" (not p-value). As expected, the ROC curve is pushed up with a conservative threshold as opposed to a more relaxed one.

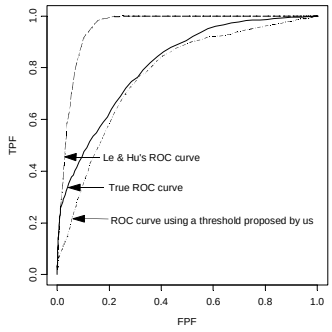

Figure 1. ROC curves from simulated data

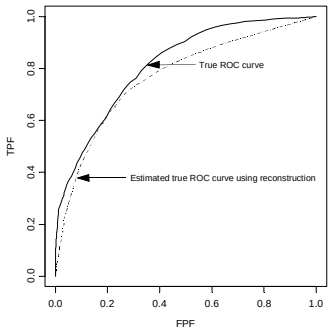

Figure 2. Comparison of true and reconstructed ROC curves

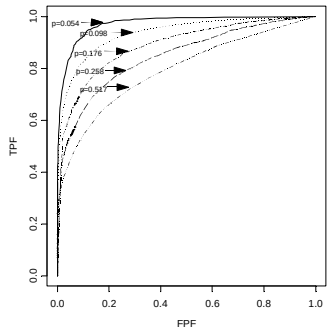

Figure 3. Le & Hu's curves for different thresholds

References

1. Sorensen J., Wang X., Magn Reson Med 1996 36: 737-744.
2. Le T. H, Hu X., NMR Biomed 1997; 10: 160-164..
3. Nandy R. R, Cordes D. Magn Reson Med 2003; 49: 1152– 1162.