Exploratory Analysis of fMR Images by Fuzzy Clustering: Voxel Preselection via "Self-Similarity"

R.L. Somorjai, M. Jarmasz, R. Baumgartner, W. Richter
Informatics Group, Institute for Biodiagnostics, NRCC, Winnipeg, MB, Canada

INTRODUCTION

Analyzing fMR images using data-driven, bias- and model-free exploratory data analysis methods, such as the Fuzzy Clustering algorithm[i] has become essential for the increasingly complex experimental designs that are devised to probe brain function. All such methods have to deal with the low level of activation above the resting state, the low SNR, and most importantly, with the fact that only a rather small fraction of the total number of voxels are activated by most stimuli. The following *preselection* method is one of several[ii] that we have devised to exclude "uninteresting" voxel time-courses (TCs) and focus only on the *potentially interesting* ones *before* carrying out a Fuzzy Clustering Analysis (FCA).

METHODS

Consider the *k*th voxel TC, $X_k(t)$, k = 1,2,...,N, defined only at discrete time points: $X_k(n)$, n = 1,2,...,T, where T is the total number of scans. We define $X_k(n)$'s *time-shifted self-similarity*, $SS_k(\ )$, as the Pearson product-moment correlation coefficient between $X_k(n)$ and $Y_k(n)$ $X_k(n+\ )$. (The sum wraps around, i.e., n + is mod(T), and is the *time shift* (lag). The wrap-around choice simplifies subsequent computations; e.g., the mean values and variances of X_k and Y_k coincide.) The preselection test is simple: for the *k*th TC we compute $SS_k(\ )$ (which may be viewed as a variant of the *serial auto-correlation function*). Suppose that there is no activation present in X_k, i.e., X_k is a TC of uncorrelated Gaussian noise. Then $SS_k(\ )$ 0 for all , and for large T it can be assumed[iii] to arise from a normal distribution with mean -1/(T-1) and variance 1/T. fMRI noise TCs are not necessarily Gaussian, nor are they uncorrelated; nevertheless, as expected, we find $SS_k(\ )$ to be small. In contrast, when a signal ("activation") is present in the TC, $SS_k(\ )$ for small (e.g., 1 or 2) is significantly different from zero. For our purposes $SS_k(1)$ (= 1) is sufficient and works well. The rationale for using $SS_k(1)$ is that it is easy and fast to compute and yet we can control statistical significance by setting a probability threshold via the standard relation between , , the Pearson correlation coefficient, and the corresponding Student's t value: $t = \{(T-2)/(1 - {}^2)\}^{1/2}$. Because $SS_k(1)$ is a correlation coefficient, it may be substituted for in the above expression. Similarly, for a given value there is a corresponding threshold $_{SS}$.

Of course, passing the SS(1) test at a given confidence level $_1$ (in EvIdent™, our exploratory data analysis software, we typically use $_1$ = 0.01) does not guarantee that the successful TC is actually "interesting"; TCs with *temporal trends* (linear or nonlinear) but no activation would also pass the SS(1) test. Therefore, prior to the SS(1) preselection step we test all TCs for trend, and *temporarily exclude* those that fail an independent statistical *trend* test, with significance $_2$. The default value in EvIdent is $_2$ = 0.05. We have implemented trend exclusion as a two-stage process. First we correlate each TC with a straight line. The TCs identified as "trendy" are then averaged to create a *trend centroid* C_{trend}. We then repeat correlating all TCs, now with C_{trend}. Consequently, the shape of C_{trend} is created by the data; it is generally highly nonlinear.

RESULTS AND DISCUSSION

For most EPI data sets with T 50, 50-70% of all brain voxel TCs seem to have trends and are excluded. (Even for FLASH data 10-30% of the TCs have trends.) From the remaining 30-50% only the potentially interesting TCs (as identified by SS(1)) are finally subjected to FCA. The remainder is placed in a "reject" cluster. It is important to emphasize that we do not permanently *remove* trends by fitting the TCs to some (typically linear) trend model (common practice in other software such as AFNI or SPM). The advantage of temporarily excluding most of the uninteresting (i.e., "trendy" and noise) TCs from analysis is not only the considerable gain in computational speed; another advantage is that these TCs no longer mask the activated ones and thus don't confound subsequent analysis. We display in the Figures an example of what TC grouping this preselection approach produces even prior to clustering. (The visual task paradigm is periodic: off-on-off-on-off, but EvIdent is not given this information.) Clearly, the expected response is already apparent (Fig.1) and is the weighted average of all TCs with $SS_k(1)$ $_{SS}$, with weights equal to $SS_k(1)$. The corresponding SS(1) map (Fig.2) strongly suggests the expected location of the activation. (In EvIdent all these can be viewed prior to clustering.) Thus FCA has an easy task in refining and further differentiating this response, and identifying additional TCs (Fig. 3). The data are from a visual EPI experiment, 64X64 with 7 slices, 61 scans. 7867 brain TCs are analyzed. Total execution time with EvIdent (including preselection and clustering) is 1.53 seconds.

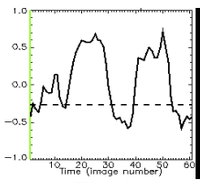

Fig. 1

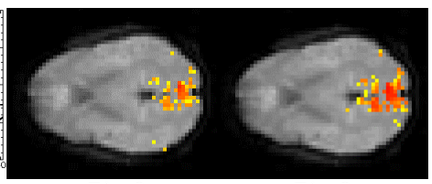

Fig. 2 **Fig. 3**

CONCLUSIONS

The preselection method we have presented is one of several useful weapons that ought to be in the arsenal of any *practically useful*, i.e., *fast* Exploratory Data Analysis (EDA) method for fMRI. It is one of two[2] preselection methods currently implemented in our EDA software EvIdent™ v. 4.31. Combined with trend exclusion[iv] and cluster merging[v], the execution speed of EvIdent's FCA approaches real-time: data sets ranging in size between 2 and 135 Mbytes have been analyzed in 0.5 – 21 secs on our SGI R10000 180 MHz CPU (which is somewhat slower than a 400 MHz Pentium II PC).

REFERENCES

[i] Scarth, G. & al. ISMRM 1995, Nice, France, p. 238.
[ii] Jarmasz, M. & Somorjai, R.L. ISMRM 1999, Philadelphia (submitted)
[iii] Madansky, A. *Prescriptions for working statisticians*, Springer-Verlag, New York (1988)
[iv] Somorjai, R.L. & Jarmasz, M. ISMRM 1999, Philadelphia (submitted)
[v] Jarmasz, M. & Somorjai, R.L. ISMRM 1998, Sydney, Aus., p. 2068