ROBUST CORRELATION ANALYSIS WITH AN APPLICATION TO FUNCTIONAL MRI

Joakim Rydell, Magnus Borga, Hans Knutsson

Dept. of Biomedical Engineering and
Center for Medical Image Science and Visualization,
Linköping University

ABSTRACT

Correlation is often used to measure the similarity between signals and is an important tool in signal and image processing. In some applications it is common that signals are corrupted by local bursts of noise. This adversely affects the performance of signal recognition algorithms. This paper presents a novel correlation estimator, which is robust to locally corrupted signals. The estimator is generalized to multivariate correlation analysis (general linear model, GLM, and canonical correlation analysis, CCA). Synthetic functional MRI data is used to demonstrate the estimator, and its robustness is shown to increase the performance of signal detection.

***Index Terms*—** Correlation, Robustness, Biomedical image processing, Magnetic resonance imaging

1. INTRODUCTION

Correlation is used in many different applications where similarities between signals are measured. Examples include speech recognition [7], fingerprint recognition [1], evaluation of image registration algorithms [8] and analysis of functional MRI (fMRI) data [4]. In many of these cases, the correlation between a set of known reference signals and a newly recorded signal is calculated in order to classify the new signal as either belonging to the same group as the reference signals or not. In the biometric applications, the correlation is used to determine whether for example a speaker is a certain person or not, while in the fMRI context it is used to determine if a certain part of the brain is activated. The reference signals may be recorded in advance or calculated from a model. In either case, it is often possible to ensure a certain quality of these signals since they are recorded or calculated under controlled circumstances. The situation is worse for the new signals, which are compared to the references. These entire signals may have a low signal-to-noise ratio, or sometimes most of the signals are well represented by the recording, but some segments are corrupted by for example background noise or radio frequent interference. Since these signals are recorded while a system is being used, it is often not acceptable to re-record until signals of good quality are obtained. In a voice recognition system re-recording translates to making users repeat spoken commands, and in the fMRI application it translates to repeating an entire experiment. The latter is both troublesome for the patient and expensive. Hence a way to measure correlations robustly, i.e. in such a way that corrupted signal recordings have minimal influence on the correlation estimates, is desirable.

This paper presents an estimator of correlation which is robust to corrupted segments in the signals. The paper is organized as follows: in the next section, the theory of correlation and its generalizations to multidimensional signals (the general linear model and canonical correlation analysis) is reviewed. In section 3 the proposed method is explained. Section 4 presents how the method can be applied to fMRI data analysis. A discussion of the method is presented in section 5, and finally section 6 concludes the paper.

Thanks to the Swedish Research Council, SMIV and MOVIII for funding.

2. THEORY

2.1. Correlation, GLM and CCA

The ordinary Pearson correlation ρ between two one-dimensional signals x and y is defined as the covariance of the signals divided by the geometric mean of their respective variances, i.e.

$$\rho = \text{Corr}(x, y) = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)\text{Var}(y)}} \tag{1}$$

Assuming that both x and y are sampled signals of length N with zero mean, an estimate of the correlation, $\tilde{\rho}$, can be calculated as

$$\tilde{\rho} = \frac{\sum_{i=1}^{N} x_i y_i}{\sqrt{\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i^2}} \tag{2}$$

If the variables are not zero-mean, their respective mean values are simply subtracted prior to this calculation.

The concept of correlation can be extended to multidimensional variables. The first extension is named the general linear model (GLM) and handles correlations between a one-dimensional and a multidimensional variable. The correlation is then defined as the maximum correlation (disregarding the sign) between the one-dimensional variable and any one-dimensional projection of the multidimensional variable, i.e.

$$|\rho| = \max_{\mathbf{w}} |\text{Corr}(x, \mathbf{w}^T \mathbf{y})| \tag{3}$$

where the vector $\mathbf{w}$ defines the projection of $\mathbf{y}$ which maximizes the correlation with x.

A further generalization is termed canonical correlation analysis (CCA) [6] and handles correlations between two multidimensional variables. The canonical correlation is defined as the maximum correlation (still disregarding the sign) between any one-dimensional projections of the two variables, i.e.

$$|\rho| = \max_{\mathbf{w}_x, \mathbf{w}_y} |\text{Corr}(\mathbf{w}_x^T \mathbf{x}, \mathbf{w}_y^T \mathbf{y})| \tag{4}$$

In CCA, the estimated correlation $\tilde{\rho}$ can be written as

$$\tilde{\rho} = \frac{\mathbf{w}_x^T \left(\sum_{i=1}^N \mathbf{x}_i \mathbf{y}_i^T\right) \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T \left(\sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i^T\right) \mathbf{w}_x \mathbf{w}_y^T \left(\sum_{i=1}^N \mathbf{y}_i \mathbf{y}_i^T\right) \mathbf{w}_y}} = \qquad (5)$$

$$= \frac{\mathbf{w}_x^T \mathbf{C}_{xy} \mathbf{w}_y}{\sqrt{\mathbf{w}_x^T \mathbf{C}_{xx} \mathbf{w}_x \mathbf{w}_y^T \mathbf{C}_{yy} \mathbf{w}_y}}$$

where $\mathbf{C}_{xx}$ is the within-set covariance matrix of the multidimensional variable $\mathbf{x}$, $\mathbf{C}_{xy}$ is the between-sets covariance matrix of $\mathbf{x}$ and $\mathbf{y}$ etc. The canonical correlation can be calculated by solving the eigenvalue problem

$$\begin{pmatrix} \mathbf{C}_{xx} & 0 \\ 0 & \mathbf{C}_{yy} \end{pmatrix}^{-1} \begin{pmatrix} 0 & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{w}_x \\ \mathbf{w}_y \end{pmatrix} = \rho \begin{pmatrix} \mathbf{w}_x \\ \mathbf{w}_y \end{pmatrix} \qquad (6)$$

The first eigenvalue is the estimate of the canonical correlation, while the first eigenvector is the concatenation of the corresponding projection directions $\mathbf{w}_x$ and $\mathbf{w}_y$.

Obviously, GLM is obtained as a special case of CCA when either $\mathbf{x}$ or $\mathbf{y}$ is one-dimensional. Ordinary correlation is the special case obtained when both variables are one-dimensional.

2.2. Weighted correlation

A weight can be assigned to each sample, allowing different samples to affect the correlation estimate to a different extent. This is useful if each sample is accompanied by a certainty value, to make less certain samples affect the correlation estimate less than more certain ones. Assuming that c_i is the certainty (or weight) associated with the i:th sample, the weighted equivalent of equation 2 becomes

$$\tilde{\rho} = \frac{\sum_{i=1}^N c_i x_i y_i}{\sqrt{\sum_{i=1}^N c_i x_i^2 \sum_{i=1}^N c_i y_i^2}} \qquad (7)$$

It is important, however, to consider the mean values of the signals. Before introducing weights, we could simply subtract the averages to obtain zero mean signals. Since samples with low weights should not affect the correlation estimates to a large extent, this is no longer the case. Instead, the weighted average must be subtracted before applying equation 7. That is, the correlation should be estimated as

$$\tilde{\rho} = \frac{\sum_{i=1}^N c_i (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^N c_i (x_i - \mu_x)^2 \sum_{i=1}^N c_i (y_i - \mu_y)^2}} \qquad (8)$$

where

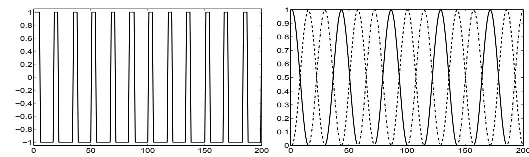
$$\mu_x = \frac{\sum_{i=1}^N c_i x_i}{\sum_{i=1}^N c_i} \qquad (9)$$

and μ_y is calculated equivalently.

GLM and CCA are adapted in a similar fashion to accommodate weighted calculation of correlation coefficients.

3. METHOD

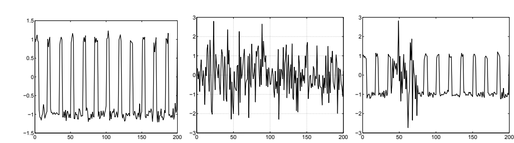
In order to use weighted correlation, the weights of all samples must be known. In some cases the weights may be available beforehand, for example if there is a natural way to measure the certainty of each sample. This may be the case if e.g. a large number of reference signals is available, since the standard deviation of each sample may be calculated and its inverse used as certainty. More often, however, no natural certainty estimate is available. Fortunately, it is possible to find suitable weights using only the two signals whose correlation is to be calculated, by automatically finding outliers in the signals. This is accomplished by dividing both signals into a number of segments. The correlation between the signals is then calculated in each segment using either Pearson correlation, GLM or CCA depending on the dimensionality of the signals. By the use of weighted correlation the segments need not be disjoint subsets of the signals, but can instead be defined by partly overlapping smooth windows. This is preferable to disjoint subsets since it makes the algorithm less sensitive to the exact locations of the boundaries between different segments. Also, it makes a smooth final weighting of the samples possible. Naturally, it is important to make sure that the sum of the weights for all windows is constant for all samples. Figure 1 shows a one-dimensional signal and a set of functions defining weights for each segment of the signal. In this example a truncated $\cos^2$ function is used to define each window and consecutive windows overlap by 50%. More overlap may be desired depending on the specific application.



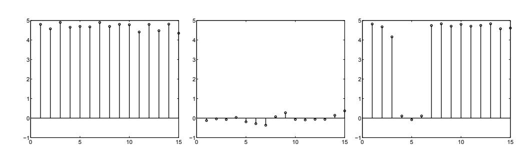
(a) Example signal (b) Window functions

Fig. 1. An example signal and a set of possible weighting functions. The windows are defined by truncated $\cos^2$ functions and overlap by 50%.

If we calculate the weighted correlation estimates between this signal and each of the signals shown in figure 2a-c, using the windows shown above, a number of correlation coefficients are obtained. These are remapped to the similarity values shown in figure 2d-f. Obviously, the similarity is approximately the same in all windows for the slightly noisy signal shown in figure 2a, and approximately zero in all windows for the noise signal in figure 2b. For the signal in figure 2c, however, the similarity is significantly lower in the noisy segment than in the other segments. The ordinary, unweighted, correlation coefficients for these entire signals are 0.99, 0.04 and 0.73, respectively.



(a) Signal 1 (b) Signal 2 (c) Signal 3



(d) Similarity 1 (e) Similarity 2 (f) Similarity 3

Fig. 2. Three signals and their similarities (in each window) to the signal shown in figure 1a.

Let us now focus on the third signal. We have a sequence of similarities, one for each window. These values can easily be split into two groups: all the relatively high values and the three low values. Since there are far more high values than there are low ones, it is intuitively obvious that the low values could be considered as outliers and that the corresponding segments of the signal should have lower weights. The algorithm arrives at the same conclusion by calculating the differences between the similarity values of every pair of signal segments. The weight of each segment is then defined as a decreasing function of the average distance to the similarity values of all other segments. More formally w_k, the weight of the k:th segment, is calculated as

$$w_k = e^{d_k^2/(2\sigma^2)} \tag{10}$$

where

$$d_k = \frac{1}{L-1} \sum_{j \neq k} |\Lambda(\tilde{\rho}_k) - \Lambda(\tilde{\rho}_j)| \tag{11}$$

and

$$\Lambda(\rho) = \frac{\text{sign}(\rho)}{1 - \rho^2 + r} \tag{12}$$

L is the number of segments. The function $\Lambda(\rho)$, which is similar to Wilks' lambda, is used to map the estimated correlation coefficients to similarity values in order to obtain a more linear scale. That is, Λ compensates for the fact that a small increase of the resemblance between two signals with low original correlation increases the correlation more than a small increase of the resemblance between two signals with high original correlation. r is a regularization parameter which controls the behavior of Λ when ρ approaches 1.

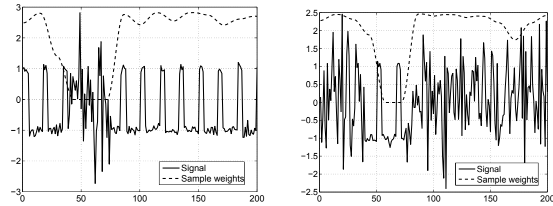
Finally, the weight of each sample can be calculated as the sum of all window functions at the sample position, multiplied by the weight for the respective segments of the signal. That is, the weight c_i of the i:th sample is calculated as

$$c_i = \sum_{j=1}^{L} w_j f_j(i) \tag{13}$$

where $f_j(i)$ is the value of the j:th window function at the position of the i:th sample. Thus for the signal in figure 2c, we end up with the weighting shown in figure 3a. The weighted correlation is 0.99, which is very close to the (unweighted) correlation between the first signal above and the reference signal. The automatically calculated weights cause the weighted correlation estimate to disregard the outliers. If the method instead is applied to any of the two other signals, an almost constant weighting is obtained. Thus, in those cases the weighted correlation estimate is the same as the ordinary correlation, which is what we want for signals whose correlation to the reference signal is constant.

Figure 3b shows a fourth example of a signal whose correlation to the reference signal in figure 1a is to be calculated. This signal consists of only noise, except for a short segment where it is similar to the reference signal. The weighting obtained is also shown in the figure. The robust correlation between this signal and the reference is 0.03, which is very close to the correlation between the reference and the pure noise signal above. Together, the third and fourth examples illustrate how the underrepresented parts of a signal are disregarded, and how this may cause the robust correlation coefficient to be either higher or lower than the ordinary correlation.

The differences between the correlations in different segments can only be calculated according to equation 11 when the signals are one-dimensional. When the signals are multidimensional and GLM or CCA is used instead of the Pearson correlation, the projection directions $\mathbf{w}_x$ and $\mathbf{w}_y$ should also be taken into account. If the correlation coefficients in two segments are similar but $\mathbf{w}_x$ or $\mathbf{w}_y$ is different, the signals are related to each other in different ways and



(a) Final weights for the signal shown in figure 2c.

(b) Another signal and its weighting.

Fig. 3. Final weights for the signal shown in figure 2c and for another signal.

the two segments should be treated accordingly. To handle this, the distance calculation is replaced by

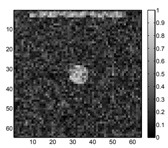
$$d_k = \frac{1}{L-1} \sum_{j \neq k} \|\Lambda(\tilde{\rho}_k)\mathbf{w}_{x_k}\mathbf{w}_{y_k}^T - \Lambda(\tilde{\rho}_j)\mathbf{w}_{x_j}\mathbf{w}_{y_j}^T\| \tag{14}$$

i.e. the difference between outer products of the projection directions (multiplied by Λ) is used to determine the distance between multidimensional correlations. The Frobenius norm is used. In the one-dimensional case, both $\mathbf{w}_x$ and $\mathbf{w}_y$ are 1, and thus equation 11 is a special case of equation 14.

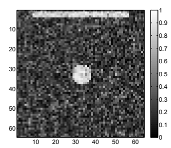
4. APPLICATION TO ANALYSIS OF FUNCTIONAL MRI DATA

Functional MRI is a technique for imaging the neural activity that arises in the brain when a task such as language processing, mental calculation or motor coordination is carried out. fMRI is based on the different magnetic properties of oxygenated and deoxygenated blood, which can be measured using magnetic resonance imaging (MRI). In short, a number of images of the brain are acquired while a subject or patient is instructed to alter between resting and performing some specific task according to a stimulus paradigm. During the active state, more oxygen is routed to those parts of the brain which contribute to the task, and this can be seen as a slightly higher intensity in the corresponding part of the images. This response to the activation stimuli is referred to as the BOLD (blood oxygen level dependent) signal. To determine what parts of the brain are activated by the task, the intensity variation over time in each voxel is examined, and the correlation to the expected BOLD signal is calculated. If the correlation in a voxel is sufficiently high, the voxel is declared to belong to a region which is activated by the task. There are two problems with this approach: the BOLD response is not exactly known and may even vary between different parts of the brain, and the acquired images are very noisy. To cope with the first problem, a subspace model of more than one BOLD signal is often used. This is where GLM is useful; a voxel should be considered as activated if its intensity variation has high correlation to any linear combination of a set of BOLD basis signals. The second problem is alleviated by spatial filtering of the images. Different methods, ranging from spatial low-pass filtering [4] to more elaborate methods for adaptive filtering [3, 9] have been proposed.

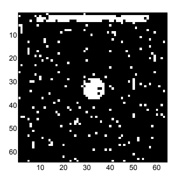
The presented estimator of correlation does not provide any improvement of the activation detection accuracy in the case where the signal to noise ratio is constant over time. If, however, the signals from some or all voxels are corrupted by sudden bursts of noise, for example induced by instruments used for delivering the stimuli to the patient or subject, the proposed method is able to automatically identify and disregard the corrupted parts of the signals. To demonstrate and evaluate the method, synthetic fMRI data was created by embedding BOLD-like signals in noise. The activated regions are a rectangle and a circle. Extra noise was added to 7 consecutive samples out of a total of 40 samples. Most often, fMRI signals are longer (approximately 100 - 200 samples), but short signals are of interest in real-time analysis of fMRI data, e.g. when using sliding window approaches [5]. Simple low-pass filtering was applied to the data before the correlation between the signal from each pixel and the BOLD model was calculated. Figure 4a and b show correlation maps obtained using ordinary correlation and the proposed method, respectively. Thresholded correlation maps are also shown. Obviously, better contrast between active and inactive pixels and fewer misclassified pixels are obtained when using the proposed estimator. A more quantitative evaluation is presented in figure 5, which shows the distributions of correlation coefficients when using ordinary correlation (a) and the proposed estimator (b). When the proposed estimator is used, the overlap between the correlation coefficients from active and inactive pixels is significantly smaller than when ordinary correlation is used. This indicates that a lower number of misclassified pixels is obtained when the proposed estimator is used.



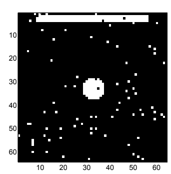
(a) Ordinary correlation



(b) Proposed estimator

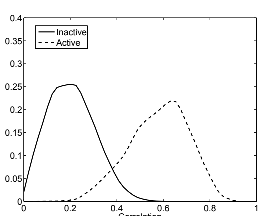


(c) Thresholded, ordinary correlation

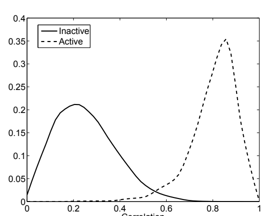


(d) Thresholded, proposed estimator

Fig. 4. Correlation maps (original and thresholded) calculated using ordinary correlation and the proposed method, respectively.



(a) Ordinary correlation



(b) Proposed estimator

Fig. 5. Distributions of correlation coefficients for active and inactive pixels.

5. DISCUSSION

Robust methods for estimating correlations have previously been proposed, see for example [2]. Most methods, however, do not consider the locality of groups of outliers. The estimator presented in this paper, on the other hand, controls the influence of each segment of the signal based on local estimates of the correlation.

For the method to work well, the correlation in the outlier segments need to be sufficiently different from the correlation in the other parts of the signal. Otherwise the outliers may not be properly detected and the weights of all segments will be similar. Then the proposed estimator will act like ordinary correlation analysis. The outlier detection is of course influenced by the choice of σ (see equation 10), but the estimator also depends on the number of segments and their amount of overlap. If the windows are too narrow, the local correlation estimates will vary greatly. On the other hand, if they are too wide, short bursts of noise may not affect the local correlation enough to be detected as an outlier. Because of this parameter dependence, the algoritm works best in applications where the properties of the signal and the noise bursts are well-known and the parameters can be fine-tuned to match the specific situation.

6. CONCLUSIONS

An estimator of correlation and its multivariate generalizations was presented. Unlike ordinary correlation estimators, the proposed algorithm is robust to locally corrupted signals. The estimator has been demonstrated on synthetic functional MRI data and it has been shown that the robustness increases the performance of signal detection. Other applications of this estimator include for example voice and fingerprint recognition as well as evaluation of image registration.

7. REFERENCES

[1] S. Angle, R. Bhagtani, and H. Chheda. Biometrics: A further echelon of security. In *UAE International Conference on Biological and Medical Physics*, 2005.

[2] C. Dehon, P. Filzmoser, and C. Croux. Robust methods for canonical correlation analysis. *Data Analysis, Classification and Related Methods*, pages 321–326, 2000.

[3] O. Friman, M. Borga, P. Lundberg, and H. Knutsson. Adaptive analysis of fMRI data. *NeuroImage*, 19(3):837–845, 2003.

[4] K.J. Friston, P. Jezzard, and R. Turner. Analysis of functional MRI time-series. *Human Brain Mapping*, 1:153–171, 1994.

[5] D. Gembris, J.G. Taylor, S. Schor, W. Frings, D. Suter, and S. Posse. Functional magnetic resonance imaging in real time (FIRE): Sliding-window correlation analysis and reference-vector optimization. *Magnetic Resonance in Medicine*, 43(2):259–268, 2000.

[6] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28:321–377, 1936.

[7] M. J. Hunt. Spectral signal processing for ASR. In *Proc. IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, 1999.

[8] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes. A comparison of similarity measures for use in 2-d-3-d medical image registration. *IEEE Transactions on Medical Imaging*, 17:586 – 595, 1998.

[9] J. Rydell, H. Knutsson, and M. Borga. On rotational invariance in adaptive spatial filtering of fMRI data. *NeuroImage*, 30(1):144–150, March 2006.