Sensor-level Maps with the Kernel Two-Sample Test

***Abstract*—Traditional approaches to create sensor-level maps from magnetoencephalographic (MEG) data rely on mass-univariate methods. In order to overcome some limitations of univariate approaches, multivariate approaches have been widely investigated, mostly based on the paradigm of classification. Recently a multivariate two-sample test called kernel two-sample test (KTST) has been proposed as an alternative to classification-based methods. Unfortunately the KTST lacks methods for neuroscientific interpretation of its result, e.g. in terms of sensor-level maps. In this work, we address this issue and we propose a cluster-based permutation kernel two-sample test (CBPKTST) to create sensor-level maps. Moreover we propose a procedure that massively reduces the computation so that maps can be produced in minutes. We report preliminary experiments on MEG data in which we show that the proposed procedure has much greater sensitivity than the traditional cluster-based permutation t-test.**

***Index Terms*—brain decoding ; two-sample test ; MEG ; brain maps**

I. Introduction

Multivariate pattern analysis (MVPA) [1] is often presented as an alternative to mass-univariate pattern analysis [2] in electroencephalographic (EEG) and magnetoencephalographic (MEG) data analysis tasks from neuroimaging experiments. In these experiments the neural correlates of different stimuli are studied to understand how a given mental process of interest occurs in the brain.

Mass-univariate analysis is based on performing multiple tests, one for each *univariate unit* of the data, for example one t-test for each sensor value at each timepoint. The high spatial and temporal granularity of the tests allows to interpret the results of the analysis in many useful ways, for example as sensor-level brain maps and across time. On the other side, univariate tests experience low sensitivity when the effects are distributed across space and time. Notice that when a joint inference has to be made over all univariate tests, e.g. when creating a map illustrating which sensors show significant differential activity across different stimuli, the univariate tests need a further layer of analysis to address the issue of *multiple comparisons*. The most common correction for multiple comparisons are [2]: the 1) Bonferroni correction, which assumes that all tests are statistically independent from each other, the 2) *cluster-based correction*, which assumes that effects occurs in clusters of units, and the 3) false discovery rate (FDR) correction.

Multivariate approaches frequently relies on the statistical learning framework where a classifier is built to predict the stimulus provided to the subject from concurrent MEG data. From the ability to accurately predict future stimuli, the classifier answers the question whether there is evidence about the mental process of interest within neural correlates. By analysing how the classifier uses the data, it is often possible to create brain maps. This approach has been recently criticised when compared to generative approaches [1], and there is no final consensus on how to create brain maps from classifiers.

Recently a different multivariate approach, called *kernel two-sample test* (KTST), has been proposed for the analysis of MEG data [3]. This approach recasts the multivariate discrimination problem as a high-dimensional two-sample test. The idea is to directly test whether the neural correlates of different stimuli are different. This is conceptually simpler than training a classifier and then testing its ability to correctly predict. In [3], it is empirically shown that the KTST and classifiers provide equivalent results about the discrimination problem. A major limitation for the adoption of the KTST in neuroimaging data analysis is that up to now, to the best of our knowledge, no solutions have been proposed to allow deeper interpretation of the results, for example in terms of brain maps.

In this work, we tackle the interpretability issue of the KTST and we propose a procedure to create sensor-level brain maps from MEG data. The proposed solution is based on conducting a KTST at each sensor using the timeseries of that sensor as multidimensional description of each trial. In order to cope with the multiple comparisons problem, a second level consisting of a cluster-based permutation test is used to compute the actual significance of the results. For these reasons, the proposed method is called *cluster-based permutation kernel two-sample test* (CBPKTST). Since the KTST is a permutation-based test[1], the straightforward implementation of the proposed solution would require a global permutation test of all the permutation tests conducted at sensor-level, which would require months of computation for the data of a typical MEG experiment. For this reason, another key contribution of this work is an algorithm to greatly reduce the computational burden by reusing the permutations computed at the sensor level as permutations for the global level. The proposed algorithm provides the same expected solution of the straightforward implementation and not just an approximation. The computational complexity of the proposed procedure is that of conducting one single KTST at each sensor. With the proposed approach, a sensor-level map from a typical MEG experiment can be computed in approximately 25 minutes on a standard computer. A final contribution of this work is to show preliminary empirical evidence of the greatly enhanced sensitivity of the proposed method with respect to the traditional cluster-based permutation t-test.

In the remaining part of the paper first, in Section II, we formally describe the proposed approach together with

[1]We are aware that some approximated versions of the KTST do not requite a permutation-based approach.

the cluster-based permutation t-test. Then, in Section III, we provide preliminary empirical evidence on real MEG data that the CBPKTST maps can be produced in a reasonable amount of time and that they show much greater sensitivity than the cluster-based permutation t-test. The results are discussed in Section IV together with some perspectives on future work.

II. Methods

In this section, we assume that the brain activity is represented only in the time domain, for simplicity. Additionally we assume that the stimulation protocol presents just two categories of stimulus, i.e. that the neuroscientific study is focused on a contrast of two conditions.

A. Notation

Let $X \in \mathbb{R}^{C \times Q}$, where C is the number of channels and Q the number of timepoints, be the multivariate timeseries describing the MEG brain activity recorded when a given stimulus, represented by the binary variable $y \in \mathcal{Y} = \{0, 1\}$, is presented to the subject. Let $\mathbf{x}$ be a realisation of X, then a pair $(\mathbf{x}, y)$ is called *trial*. Multiple trials are collected during and experiment and let $A = \{\mathbf{x}_i | y_i = 0, i = 1 \dots m\}$ and $B = \{\mathbf{x}_i | y_i = 1, i = 1 \dots n\}$ be two samples of trials, one for each category of stimulus. $A \cup B$ is the dataset recorded during the MEG experiment. Typical values for the quantities expressed so far, after a standard preprocessing procedure, (see Section III) are: $C \approx 300$, $Q \approx 100$ and $N = m + n \approx 500$.

B. Cluster-based Permutation Test

The cluster-based permutation test [2], [4] is based on the assumption that the effect, when present, appears in clusters of neighbouring units. Under our assumptions, a unit is a sensor at a given timepoint. Once it is defined how two units are considered neighbours, the neighbouring units exhibiting a significant activity are clustered and then a cluster-level statistic is computed. The most well-known cluster-based statistic is the *cluster-level mass*, which is defined as the sum of the unit-level statistics in each cluster. The procedure to compute the clusters and the cluster-statistic is (adapted from [2]):

1) Compute the value of test statistic T at each unit and their related p-value.
2) Keep only the units that are significant, i.e. those where p-value $< \theta$.
3) Cluster the neighbouring units.
4) Compute the cluster-level statistic for each cluster by summing the statistic value of each unit in the cluster: $T_{\text{cluster}} = \sum_{i \in \text{cluster}} T_i$.

The cluster-level statistic of each cluster, T_{cluster}, is then compared against its null distribution to get its p-value, i.e. the significance level of the cluster. The null distribution of the cluster-level mass is usually computed as the distribution of the max cluster-level mass of all clusters under the null-hypothesis [2]. In practical cases it is not possible to explicitly derive this null distribution, so it is estimated through a resampling approach, usually by drawing M random permutations of $\{y\}_{i=1..N}$ and, for each permutation, by computing the resulting clusters and cluster-based statistic. The maximum value of the cluster-based statistic over all clusters at each permutations is stored and the set of maximum values creates the estimated null distribution.

In the case of the cluster-based permutation t-test, the t-statistic is computed for each timepoint at each sensor together with the related p-value through the Student's t-distribution. Units showing significant departure from the null-hypothesis are clustered as explained in the previous procedure and the null distribution of T_{cluster} is assessed through permutations. The significance of the clusters computed on the original (non permuted) data is assessed against this null distribution and then sensor-level maps can be produced across time.

C. The Kernel Two-Sample Test

In a two-sample test problem, two samples A and B are drawn *independently and identically distributed* (i.i.d) from their respective distributions P_A and P_B. The problem is to test the hypothesis $H_0 : P_A = P_B$, based only on the information from the two samples. If H_0 is true, then the two set of trials are expected to be indistinguishable. In the decoding context, this means that the neuroimaging data, recorded during the presentation of the two kinds of stimuli, do not present systematic differences. In [5] the *maximum mean discrepancy* (MMD) is adopted as a distance between two distributions:

$$\text{MMD}[\mathcal{G}, P_A, P_B] := \sup_{g \in \mathcal{G}} (E_{x_A \sim P_A}[g(x_A)] - E_{x_B \sim P_B}[g(x_B)]). \tag{1}$$

For a specific family of kernel functions, i.e. the characteristic kernels [2], a kernel-based expression of MMD^2 can be obtained (see [5], [6] for further details):

$$\text{MMD}^2 = E[k(X_A, X'_A)] - 2E[k(X_A, X_B)] + E[k(X_B, X'_B)]. \tag{2}$$

Given two samples A and B, an unbiased estimate of MMD^2 is:

$$\text{MMD}_u^2 = \frac{1}{m(m-1)} \sum_{i \neq j} k(x_i^A, x_j^A) - \frac{2}{mn} \sum_{i,j} k(x_i^A, x_j^B) + \\ + \frac{1}{n(n-1)} \sum_{i \neq j} k(x_i^B, x_j^B) \tag{3}$$

where $A = \{x_1^A, \dots, x_m^A\}$ and $B = \{x_1^B, \dots, x_n^B\}$. In [6], this estimate is proved to converge exponentially fast to the true value.

MMD_u^2 is a multivariate test statistic that can be used for a two-sample test even on high-dimensional data because of the kernel function. The null-distribution of MMD_u^2 is problem-specific and the usual approach to its estimation is by resampling techniques. By means of M random permutations of the assignment of the trials to A and B, it is possible to create a large sample of MMD_u^2 values under H_0 and thus to estimate the p-value of MMD_u^2 for the actual dataset at hand.

[2]The well-known Gaussian/RBF and Laplacian kernels are characteristic kernels.

In this work, we propose to conduct a KTST at each sensor by using the timecourse of each trial at that sensor as its local multidimensional description. In this way, at each sensor, the MMD_u^2 distance aims at quantifying how much the neural correlates of the two stimuli systematically differ.

Notice that the value of MMD (and of MMD_u^2) is not absolute, i.e. we cannot compare the MMD values from two different problems because their underlying null distributions may be different. Thus the MMD value is problem specific and for this reason, in our case, it is necessary to compute the null distribution of MMD_u^2 at each sensor.

D. Scalable Cluster-based Permutation KTST

In this work, we propose a cluster-based permutation test based on the MMD_u^2 statistic . This approach is different from the traditional solution based on the t-test because it is multivariate: the new units are just the sensors because the time dimension is used as the multivariate description of each trial. We denote the proposed approach as *cluster-based permutation kernel two-sample test* (CBPKTST). In order to carry out a CBPKTST, two technical problems need to be solved: the inhomogeneity between MMD_u^2 values at different sensors and the excessively high computational complexity. In the following, we describe the two problems and our proposed solutions.

1) Inhomogeneity of MMD_u^2 *across sensors:* When we need to compute cluster-level statistics, we need to sum *homogeneous* unit-level statistics in order to obtain a meaningful quantity. Since the distribution of MMD_u^2 values is, in general, sensor-specific, then we need to transform the sensor-specific MMD_u^2 values into a homogeneous quantity. For each sensor/unit, we propose to use $T = 1 - p$-value, where the p-value is that of MMD_u^2 obtained through the permutation test. This is motivated by the fact that p-values - and thus $1-p$-values - are, by definition, uniformly distributed under the null hypothesis. We prefer $1 - p$-value to the p-value because we need a quantity that increases when the distance between the trials across the two categories increases.

2) Scalability: Ideally, performing a cluster-based permutation KTST is prohibitively expensive from the computational point of view. In order to compute the null distribution of the cluster-based statistic, it is necessary to compute the whole set of KTSTs, one for each sensor, for each permutation of the cluster-based statistic. This leads to M^2 permutations, each requiring the computation of C evaluations of MMD_u^2, an amount which is not feasible in practice (see Section III for actual numbers from a typical MEG experiment).

In this work, we propose a major shortcut in the computations which consists of re-using the MMD_u^2 values computed during the permutations at the unit-level. In essence, when computing KTST at each sensor, by imposing the exact same permutation sensor-wise, there is no difference between the sensor-level permuted values of the test statistic T and the values necessary to compute the cluster-level statistic, so they can be used at both levels. The proposed solution requires to store the intermediate results during the unit-level permutations and for this reason it trades-off the computational complexity with increased memory requirements. The memory required is of $C \times M$ floating-point values, which, for typical MEG experiments, is in the order of tens of megabytes (see Section III).

Here follows the detailed description of the proposed CBPKTST. Given θ, i.e. the threshold for rejecting the null-hypothesis for the test at each sensor (e.g. $\theta = 0.05$),

1) For each sensor $c = 1 \dots C$, compute the three kernel matrices $K_c^A = \{k(x_{c,i}^A, x_{c,j}^A)\}_{i,j \in 1 \dots m}$, $K_c^B = \{k(x_{c,i}^B, x_{c,j}^B)\}_{i,j \in 1 \dots n}$, $K^{AB} = \{k(x_{c,i}^A, x_{c,j}^B)\}_{i \in 1 \dots m, j \in 1 \dots n}$.
2) Compute $\mathrm{MMD}_u^2(c)$ for each sensor following Eq.3.
3) Do a permutation test for each sensor and store each permuted value of MMD_u^2 in a $C \times M$ matrix, $\mathrm{MMD}_u^2[c, l]$, $c = 1 \dots C$, $l = 1 \dots M$ [3]. Notice that the exact same permutation has to be done across all sensors at each iteration.
4) Compute the test statistic $T = 1 - p$-value for each $\mathrm{MMD}_u^2[c, l]$, by sorting the rows of the matrix.
5) For each permutation:
   a) Compute the spatial clusters in sensor space among the sensors which have a sufficiently low p-value, i.e. p-value $\leq \theta$.
   b) Compute and store the resulting max cluster-statistic $\mathcal{T} = \max_{\gamma \in \text{clusters}} \sum_{\text{sensor} \in \gamma} T_{\text{sensor}}$.

The computational bottleneck of this procedure is Step 3, which requires $M \times C$ evaluations of Equation 3, i.e. $M \times C \times N^2$ sums. Notice that each evaluation is independent and thus this step can be easily split into parallel computations.

A free / open source reference implementation of the proposed algorithm is provided in Python at https://github.com/***ANONYMISED***.

III. Experiments

We compared the proposed CBPKTST described in Section II-D against the cluster-based t-test described in Section II-B on data from a real MEG experiment that we collected for a forthcoming study. The stimulation protocol comprised the following visual stimuli: Face, House and Body, with a balanced design. The visual stimuli were shown either on the left or on the right side of the screen while the subject stared at a fixation cross. Moreover the face stimulus was half of the time a male face and the other half a female face. Five different two-sample problems were studied: left vs. right, face vs. house, face vs. body, house vs. body and female vs. male. The resulting sensor maps, where significant clusters were detected, are presented in Figure 1. For lack of space, the results we present here are about only one subject of the study.

The MEG data were collected with an Elekta Neuromag scanner comprising 102+102 gradiometers and 102 magnetometers. Then 306 timeseries, each of 3s and sampled at

[3] In order to compute the permuted values of MMD_u^2 we just need to rearrange rows and columns of K_c^A, K_c^B and K_c^{AB}, so no further evaluations of $k(\cdot, \cdot)$ are necessary.

1KHz, were recorded for each of the 677 trials of the experiment. The pre-processing of the recorded signal was: 1Hz high-pass filter, baseline removal (baseline: average signal in $[-1s, 0s]$), downsampling to 100Hz, and trimming each trial to $[0s, 1s]$. Gradiometers were merged in pairs and magnetometer data was removed. For each trial, the 102 preprocessed timeseries were concatenated into a single vector of 10200 values. For each of the five two-sample problems, a dataset was created keeping just the trial relevant for each problem. Each dataset was then z-scored with the grand mean and grand standard deviation.

In order to define proximity between sensors and between timepoints, as required by the cluster-based assumption of both tests, we followed the indications in [2]: 5.4cm between sensors and temporal distance of 1 timepoint. For the CBPKTST we adopted the Gaussian kernel with the σ^2 estimated as the squared median distance between al pairs of trials.

The threshold for sensor-level and cluster-level tests was set to $\theta = 0.05$. The number of permutations was set to $M = 10000$. For each two-sample problem the time to compute the cluster-based permutation t-test was ≈ 10 minutes and for the CBPKTST it required ≈ 25 minutes on a quad-core 2.4GHz Intel CPU. The memory requirement for CBPKTST was $\approx$ 40Mb. As a comparison, the estimated time required to run the straightforward implementation of CBPKTST would have been M times more, i.e. 6 months.

Additionally, we note the same dataset was used in [ANONYMISED] and showed that for the first 4 of the 5 contrasts, i.e. left vs. right, face vs. house, face vs. body and house vs. body, accurate classification was achieved. These previous results support the claim that differential brain activity is present in the data and that evidence of that should be detected also in sensor-level maps.

IV. Discussion & Conclusion

In this work, we proposed a multivariate procedure for creating sensor-level maps from single-subject MEG data of neuroimaging experiments. The proposed procedure is based on the KTST applied at each sensor and a further layer of cluster-based permutation test to account for spatial correlations and multiple testing. The proposed procedure can be seen as a multivariate extension of the cluster-based permutation t-test, which is univariate.

The proposed efficient implementation of the CBPKTST reduces the cost of the computation by several order of magnitude with respect to straightforward implementation, i.e. from months to minutes for data of a typical MEG experiment.

Preliminary results on MEG data shows that the proposed approach confirms the results in [ANONYMISED] and finds significant clusters in occipital, parietal and temporal areas (see Table 1). On the contrary, the cluster-based permutation t-test is not able to find any significant cluster of activity within the same data, for each contrast. This evidence support the claim that the sensitivity of the proposed method is much superior than that of the cluster-based permutation t-test.

The proposed approach is able to provide means for interpretation of the spatial patterns of the mental activity associated to the contrasts of interest. Conversely, by exploiting the multivariate aspect of the signal in time, there is no interpretation over the time axis. For this reason, future work will address the use of MMD as a multivariate distance for other multivariate representations of MEG data.

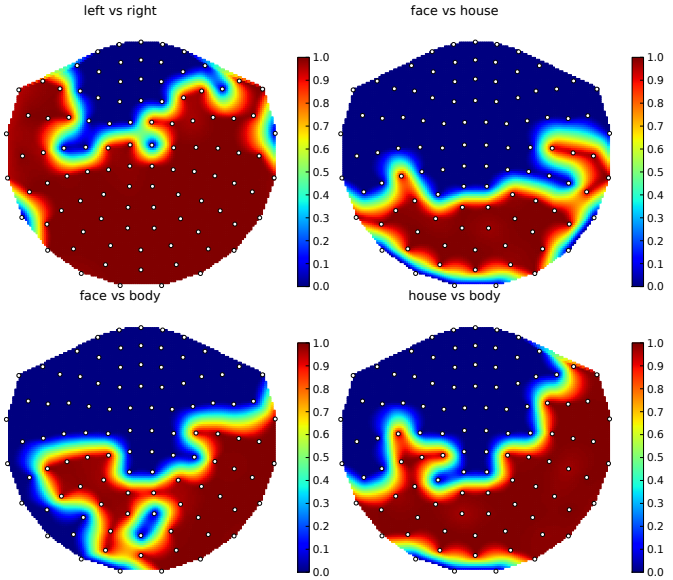

Fig. 1. Sensor-level (gradiometers) brain maps obtained with the CBPKTST ($M = 10000$, $\theta = 0.05$). The color indicates the homogeneous statistic at sensor-level. The contrast is indicated on top of each map. Only significant clusters are shown, in red. The contrast *female vs male* had no significant clusters so it is not reported. The cluster-based permutation t-test had no significant clusters on all contrasts, so the related maps are not reported.

References

[1] S. Haufe, F. Meinecke, K. Görgen, S. Dähne, J.-D. Haynes, B. Blankertz, and F. Bießmann, "On the interpretation of weight vectors of linear models in multivariate neuroimaging," *NeuroImage*, vol. 87, pp. 96–110, Feb. 2014. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2013.10.067

[2] D. M. Groppe, T. P. Urbach, and M. Kutas, "Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review." *Psychophysiology*, vol. 48, no. 12, pp. 1711–1725, Dec. 2011. [Online]. Available: http://dx.doi.org/10.1111/j.1469-8986.2011.01273.x

[3] E. Olivetti, D. Benozzo, S. M. Kia, M. Ellero, and T. Hartmann, "The Kernel Two-Sample Test vs. Brain Decoding," in *Proceedings of the 2013 International Workshop on Pattern Recognition in Neuroimaging*, ser. PRNI '13. Washington, DC, USA: IEEE Computer Society, Jun. 2013, pp. 128–131. [Online]. Available: http://dx.doi.org/10.1109/prni.2013.41

[4] E. T. Bullmore, J. Suckling, S. Overmeyer, S. Rabe-Hesketh, E. Taylor, and M. J. Brammer, "Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain." *IEEE transactions on medical imaging*, vol. 18, no. 1, pp. 32–42, Jan. 1999. [Online]. Available: http://dx.doi.org/10.1109/42.750253

[5] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola, "A Kernel Method for the Two-Sample-Problem," in *NIPS*, 2006. [Online]. Available: http://books.nips.cc/papers/files/nips19/NIPS2006_0583.pdf

[6] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A Kernel Two-Sample Test," *Journal of Machine Learning Research*, 2012. [Online]. Available: http://www.jmlr.org/papers/volume13/gretton12a/gretton12a.pdf