Tackling the cognitive processes that underlie brands’ assessments using artificial neural networks and whole brain fMRI acquisitions

José Paulo Santos 1, 2
1 ISMAI, Maia, Portugal
2 Socius - Research Centre in Economic and Organizational Sociology, Lisbon, Portugal
jpsantos@ismai.pt

Luiz Moutinho 3
3 Business School, Department of Management,
University of Glasgow, Glasgow, Scotland
l.moutinho@bss.gla.ac.uk

Abstract — This exploratory study proposes the use of artificial neural networks to analyze whole brain fMRI data. Because fMRI data is dimensionally exorbitant, the first step is to reduce the amount of data to a tractable size, which is accomplished using probabilistic independent component analysis (PICA). Then data enters a simple backpropagation feedforward neural network. This network outputs correct predictions above chance level in a different sample of subjects. More interestingly, it is found that hidden nodes segregate and concentrate different, but coherent, brain networks, which are the target of interpretations to support cognitive processes during the assessment of brands’ logos.

Keywords—fMRI; artificial neural networks; marketing; brand; evaluation

I. INTRODUCTION

In order to surpass recognized limitations of conventional mass univariate fMRI data analysis, other methods that do not assume voxels’ inter-independence have been proposed [1, 2]. Hanson et al. [3] proposed the use of artificial neural networks (ANNs) with fMRI data to challenge the established assumption that category representation in the brain is localized. Instead, they concluded that category representation is distributed and has a combinatorial nature, failing the surfacing of single category representations. Thus, their study proved that it is possible to use ANNs to analyze fMRI data, and that such method may be more thoughtful and judicious than univariate approaches when rising conclusions.

However, the use of multivariate methods like ANNs introduces vast complexity and computing load into the analysis. Because of this, as in the study of Hanson et al., such methods have been applied only to limited parts of the brain and such compromise has been suggested [4]. Nonetheless, a whole brain analysis remains an interesting objective for cognitive sciences.

Misaki and Miyachi [5] also explored the use of ANNs to model the blood-oxygen-level dependent (BOLD) signal that is the basis of functional magnetic resonance. Their approach is the inverse of the one proposed in the presented study because the BOLD signal is used to supervise the network training. They conclude that ANNs are suitable for BOLD signal analysis and suggest strategies to overcome autocorrelation noise.

Sona et al. [6] also used ANNs to predict stimuli categories from fMRI signal, and concluded that ANNs, especially recurrent neural networks, perform better than traditional GLM analysis. To reduce computing load, they restricted the analysis to voxels’ clusters that hold information for classification purposes. The clustering procedure included a spatial and a temporal dimension. Similarly Espírito-Santo et al. [7] used a principal component analysis (PCA) to reduce the amount of data to a tractable size.

The present article reports an exploratory approach to design a procedure for the application of ANNs to whole brain fMRI data analysis. Current fMRI acquisitions may deliver signal for around 150,000 voxels, which constitutes an intractable dataset for ANNs. Hence, the proposed strategy is made of two stages: in the first stage dimensionality is reduced using probabilistic independent component analysis (PICA), which has been the target of scrutiny of the scientific community [8, 9]; in the second stage the independent components (ICs) enter a simple backpropagation feedforward neural network, which, after training, is used to predict brands’ assessments of a different set of subjects.

To test this procedure it is used data from a study where brands’ logos are assessed (although some of them are fictitious because were designed purposely for the experiment). This paradigm also includes a dual baseline (non-emotional words and a conventional fixation cross passive viewing), which was important for the original GLM-based analysis, but it is not required for ANNs. The analysis of the ANN focus on its performance (measuring correct predictions, i.e., the ability of the ANN to accurately guess the choices in a different sample of subjects), and also focus on hidden nodes and the ICs that each one concentrates in itself. This last analysis allows the inference of separate cognitive processes that subserve the behavioral responses in the task.

II. METHODS

A. Paradigm

The present study relies on the data reanalysis of a previous fMRI experiment. There were four different categories, plus the interstimuli interval. Each category was composed by thirty five slides (6 s each). The interstimuli interval ranged from 4 until 9 s, in 0.5 s steps. The sequence was optimized with
Three of the four categories were brands’ logos grouped in the following categories: positive, indifferent, and fictitious. The fourth category was non-emotional words. During the interstimuli interval participants fixated a cross.

Positive and indifferent brands were screened (35 of each) from an initial set of 200 logos using the PAD - pleasure, arousal, dominance scale [10-12], and the SAM - self assessment manikin [13], explained in detail elsewhere. The fictitious brands (also 35) were brands’ logos that did not exist in the market. Each logo was designed by a marketer made to resemble real ones, making them plausible for the participants. Non-emotional words, used as a forth category, were determiners, conjunctions, prepositions or adverbs.

Participants were instructed to either rate hedonically the brand (as positive, negative, indifferent, or unknown), to read covertly non-emotional words, or just to fixate a cross when required. Participants made their choices by using a response device – a button box (model Lumina LU400-PAIR; Cedrus Corporation, USA; http://www.cedrus.com).

B. Human subjects

The participants were eighteen, seven healthy male and eleven healthy female volunteers, right handed, with neither history of neurological nor psychiatric disturbances (mean age 28.2 ± 6.9 years, 19 to 41 years). This research project performed according to the Declaration of Helsinki and was approved by the local Ethics Committee.

C. Data acquisition

Functional images with axial orientation were obtained using a T2*-weighted EPI sequence in a Siemens® Magnetom Trio high field (3 Tesla) MRI scanner (Siemens AG, Germany) (TR = 3000 ms, TE = 30 ms, 64 x 64 matrix, FOV = 192 mm, 3.0 mm axial slices). The order of acquisition of the slices was interleaved, and they covered the whole brain. The study consisted in one session where 403 utile volumes were acquired. A whole brain anatomical structural scan was acquired also for each volunteer, using a T1 weighted MPRAGE protocol (256 x 256 matrix, FOV = 256 mm, 3.0 mm axial slices), for co-registration purposes. Gradient field mapping was additionally acquired for image quality control.

D. Image analysis and preprocessing

fMRI data processing was carried out using FEAT (FMRI Expert Analysis Tool) version 5.98, and also using probabilistic independent component analysis (PICA) [8] as implemented in MELODIC (Multivariate Exploratory Linear Decomposition into Independent Components) version 3.09, both part of FSL - FMRIB’s Software Library, www.fmrib.ox.ac.uk/fsl [14].

The subjects set was split in two: twelve participants were randomly assigned to the train group (five males and seven females), and the remaining six were allocated to the test group (two males and four females).

The fMRI data of the train group entered the PICA analysis for dimension reduction, which output 161 independent components (IC). The following data pre-processing was applied: masking of non-brain voxels, voxel-wise de-meaning of the data, and normalization of the voxel-wise variance. Pre-processed data were whitened and projected into a 161-dimensional subspace using probabilistic Principal Component Analysis where the number of dimensions was estimated using the Laplace approximation to the Bayesian evidence of the model order [8, 15]. The whitened observations were decomposed into sets of vectors, which describe signal variation across the temporal domain (time-courses), the subject domain and across the spatial domain (maps) by optimizing for non-Gaussian spatial source distributions using a fixed-point iteration technique [16]. Estimated component maps were divided by the standard deviation of the residual noise and thresholded by fitting a mixture model to the histogram of intensity values [8].

Logos’ assessment was performed by the means of a button box (which requires motor processing in the brain), while the non-emotional words and the fixation cross did not require the manipulation of buttons. As the aim of the present study is to understand the cognitive processes that underlie brands’ assessments, and because the last two stimuli could introduce parasitic confounds, we decided to remove them from further analysis and continue just with the logos (positive, indifferent, and fictitious).

From the pre-session (that supplied the stimuli for positive and indifferent categories), to the scanning session, subjects’ assessments varied. Although the changes in ratings were not extensive (6.3% for positive, and 32.2% for indifferent brands), only the assessments that were consistent between sessions were considered for further analysis. The same way, 12.1% of the fictitious logos were recognized, and such cases were also excluded from the analysis process. Because of this strategy the negative category does not exist in the output layer of the ANN, which is composed only by the positive, indifferent, and fictitious categories.

Features were then extracted from each of the 161 time courses. The strategy adopted was to average the second and third signals after stimulus onset. By this way, the average time distance from the onset was 5766 ± 857 ms, i.e. the signals considered were consistently in the neighborhood of the hemodynamic response peaks. At the end of this stage the result is a matrix with 949 rows (each corresponding to an epoch with the corresponding category), and 161 columns (each corresponding to one IC) plus one more column with the category code. This matrix is the training set.

The fMRI data of the test group was preprocessed in FEAT. The following pre-statistics processing was applied: motion correction using MCFLIRT [17]; slice-timing correction using Fourier-space time-series phase-shifting; non-brain removal using BET [18]; spatial smoothing using a Gaussian kernel of FWHM 5mm; grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor; highpass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma=30.0s). Registration to high-resolution structural and/or standard space images was done using FLIRT [17, 19].
The 161 brain activation maps obtained with the train group were used as masks to average the individual time courses in the test group. The same procedure for feature calculation was adopted, i.e. the second and third acquisitions after stimulus onset were averaged (average time distance from the onset was 5761 ± 872 ms, similar to the training set). Finally, the 502 epochs obtained were normalized for each subject. At the end of this stage the result is a similar matrix with 994 rows and 161 columns (each corresponding to one IC) plus one more column with the category code (that is used to assess the ANN calculations). This matrix is the testing set.

E. Parameters of the artificial neural networks

The AMORE package [20] implemented in R [21] was used to design and perform the necessary calculations of the backpropagation feedforward artificial neural network. Exploratory analyses yielded a global learning rate of 0.030 and a global momentum of 0.5. Similarly, a hidden layer with four nodes produced the best results. The selected activation function for the hidden nodes was “tansig”, while for output neurons the function was “sigmoid”.

In order to investigate possible bias derived from the network structure, the ANN was also fed twice with matrices similar to the test set, but now including random values from two distributions: uniform and normal.

III. RESULTS

The PICA analysis returned 161 independent components, which account with 85.86% of the variability.

The results of the ANN that performed better (largest accuracy) are represented in Table 1.

Table 1Confusion matrix with the global results of the test stage, with three sets of data: the original data set kept for the test stage (column hits), and random values from uniform and normal distributions (columns random uniform and random normal).

<table>
<thead>
<tr>
<th>Category</th>
<th>Predicted assessment</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>Fictitious</td>
</tr>
<tr>
<td>Real ass.</td>
<td>161</td>
<td>15</td>
</tr>
<tr>
<td>Fictitious</td>
<td>31</td>
<td>108</td>
</tr>
<tr>
<td>Indifferent</td>
<td>29</td>
<td>57</td>
</tr>
<tr>
<td>Total</td>
<td>221</td>
<td>180</td>
</tr>
</tbody>
</table>

The global accuracy of the ANN is 64.9% and the balanced accuracy (because the categories are somewhat unbalanced) is 63.1%. Table 2 reports the partial accuracies and precisions for each category. The K-hat (Cohen) is 0.465.

Table 2 Accuracies and precisions of the predictions for the three categories.

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Fictitious</th>
<th>Indifferent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>84.7%</td>
<td>63.9%</td>
<td>39.9%</td>
</tr>
<tr>
<td>Precision</td>
<td>72.9%</td>
<td>60.0%</td>
<td>56.4%</td>
</tr>
</tbody>
</table>

Table 3 contains the results of the classification performance of the ANN when fed with random values from a uniform and normal distribution. There is a drop in the global accuracy (33.8% for uniform and 32.7% for normal distribution), as well for the balanced accuracy (32.9% for uniform and 31.8% for normal distribution), and also for the K-hat (Cohen), which is -0.013 for the uniform and -0.027 for the normal distribution.

Table 3 Confusion matrix with the global results of the test stage with random values from an uniform (top row in each cell) and normal (bottom row) distributions.

<table>
<thead>
<tr>
<th>Category</th>
<th>Predicted assessment</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>Fictitious</td>
</tr>
<tr>
<td>Real ass.</td>
<td>79</td>
<td>82</td>
</tr>
<tr>
<td>Fictitious</td>
<td>72</td>
<td>63</td>
</tr>
<tr>
<td>Indifferent</td>
<td>65</td>
<td>54</td>
</tr>
<tr>
<td>Total</td>
<td>213</td>
<td>199</td>
</tr>
</tbody>
</table>

Table 4 lists the weights of the axons that link hidden nodes to the output nodes, while Table 5 lists the weights of selected input nodes that have important contributions for the hidden nodes.

Table 4 Weights of the axons that link hidden nodes to output nodes.

<table>
<thead>
<tr>
<th>Output nodes</th>
<th>Hidden nodes</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>-2.840</td>
<td>1.602</td>
<td>-3.282</td>
<td>-1.744</td>
<td></td>
</tr>
<tr>
<td>Fictitious</td>
<td>-2.317</td>
<td>1.454</td>
<td>-1.467</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indifferent</td>
<td>1.728</td>
<td>1.764</td>
<td>3.623</td>
<td>3.022</td>
<td></td>
</tr>
</tbody>
</table>

Table 5 Weights of the axons that link selected input nodes to hidden nodes.

<table>
<thead>
<tr>
<th>Input nodes</th>
<th>Hidden nodes</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC 10</td>
<td>-3.837</td>
<td>-3.676</td>
<td>7.816</td>
<td>3.170</td>
<td></td>
</tr>
<tr>
<td>IC 19</td>
<td>5.434</td>
<td>3.589</td>
<td>-7.333</td>
<td>-3.030</td>
<td></td>
</tr>
<tr>
<td>IC 25</td>
<td>1.733</td>
<td>0.879</td>
<td>0.299</td>
<td>0.695</td>
<td></td>
</tr>
<tr>
<td>IC 45</td>
<td>1.426</td>
<td>1.099</td>
<td>-1.789</td>
<td>-0.998</td>
<td></td>
</tr>
<tr>
<td>IC 68</td>
<td>2.258</td>
<td>1.042</td>
<td>-0.007</td>
<td>-0.602</td>
<td></td>
</tr>
</tbody>
</table>

IV. DISCUSSION

The analysis of Table 1 leads to the conclusion that the ANN holds important information for correct classification. This is because correct predictions are always higher than random values.

IC 19 has the highest weights for hidden nodes 1 and 2 and the most negative weights for hidden nodes 3 and 4, while IC 10 is the opposite. Both IC 19 and 10 extensively include brain motor regions. These results suggest that motor processes are very strong and may outshine more subtle cognitive processes that may be more informative to understand decision-making.

Hidden nodes 1 and 2 have important contributions to positive classifications and, with a lesser extent, to indifferent
brands. IC 68 (which encompasses the ventro medial prefrontal cortex - vmPFC - and the frontal pole), and IC 25 (which includes the posterior cingulate gyrus and the precuneus) have important contributions for these nodes. The vmPFC is a brain region that previously was found to participate in brand preference [22], and the precuneus was found to support autobiographical memories [23]. IC 45 represents another brain network important for positive brands classification, and extensively overlaps the Default Network, which has been found to support self-reflexive processes [24, 25].

It is then possible to find a coherent concentration of brain networks around hidden nodes 1 and 2 that support positive brands’ assessments. Each one brings relevant cognitive processes for preference decision-making (e.g. autobiographical memories and self-relatedness). The ANN extracts from these brain networks pertinent information to compute correct predictions much above chance.

An extensive analysis of the participation of the input nodes in the hidden layer is not possible due to space constrictions. However, it is interesting to note that predicting indifferent brands is considerably less effective than predicting positive or even fictitious brands, which may suggest pertinent interpretations for the Marketing discipline.

V. CONCLUSIONS, AND LIMITATIONS

The application of ANNs to whole brain fMRI data is possible. As demonstrated, the use of ANNs can model complex cognitive processes like brands’ assessments. On the one hand this model can predict choices above chance level. On the other hand, very importantly, the hidden nodes organize into separated and sounding cognitive processes. This opens the possibility to define cognition, not based on explicit task outcomes, but relying on implicit neural substrates.

Motor components were systematically present with high weights. This is because the experiment used a button box device to record choices, which is a common procedure in neuroeconomic studies. It can be argued that these parasitic processes may attract, inflate and drag other components, biasing the results. However, the lack of physical recordings is also a difficult hurdle; for example, indifferent brands’ rates were observed to be elusive, changing between sessions, which caution the use of a posteriori questionnaires.

REFERENCES