ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.

Designing Semantic Feature Spaces for Brain-Reading

L. Pipanmaekaporn[1]*, L. Tajtelbom[2], V. Guigue[2] and T. Artières[3] †

1- King Mongkut's University of Technology North Bangkok, Thailand

2- Laboratoire d'Informatique de Paris 6 (LIP6) Paris, France

3- Laboratoire d'Informatique Fondamentale (LIF) Marseille, France.

Abstract.

We focus on a brain-reading task which consists in discovering a word a person is thinking of based on an fMRI image of their brain. Previous studies have demonstrated the feasibility of this brain-reading task through the design of what has been called a semantic space, i.e. a continuous low dimensional space reflecting the similarity between words. So far the best results have been achieved by carefully designing this semantic space by hand which limits the generalization of such a method. We propose to automatically design several semantic spaces from linguistic resources and to combine them in a principled way and achieve results comparable to that of manually built semantic spaces.

1 Introduction

Neuroimaging has gained much interest in the last decade in many fields ranging from philosophy and psychology to neuroscience and artificial intelligence. Among brain imaging techniques, functional Magnetic Resonance Imaging (fMRI) has become a primary tool to detect mental activity with great spatial resolution [1]: an fMRI image contains approximately 20,000 voxels (volumantic pixels) that are activated when a human performs a particular cognitive function (e.g., reading, mental imagery) [2]. With fMRI, it became possible to associate brain areas with cognitive states: specific conceptual words and pictures trigger specific activity in some parts of the brain and studies began to focus on the extraction of meaningful brain activation patterns [3, 4].

A pioneering work [5] showed that it was possible to predict the brain activation pattern (a fMRI image) in response to a given conceptual stimulus (e.g. a word). Reciprocally [6] demonstrated on the same dataset the feasibility of identifying the concept from the brain activation pattern (fMRI image). The proposed approaches for these two reciprocal tasks share the definition of an intermediate semantic (or representation) space to represent the concepts, the underlying idea being that it allows the problem of inferring the concept from the fMRI (and vice versa) to come down to a standard regression problem from the fMRI voxel space to the semantic space (and vice versa). Importantly if

*This work was done while the author was at LIP6 for an internship.

†We want to thank the French government for having funded the stay of L. Pipanmaekaporn in France within the Franco-Thai Junior Research Fellowship program from May to July 2014.

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.

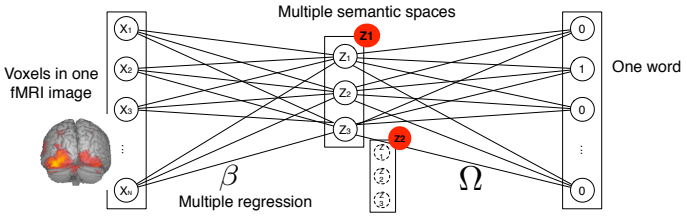

Fig. 1: Brain-reading processing chain

the representation space is designed in such a way that one can get the representation of any new word, such a strategy naturally allows the recognition of concepts from an fMRI image even if there was no training fMRI image for this word. This may be done in two steps: first, starting from an input fMRI image, a point in the semantic space is computed using the regression model; then, the word whose representation is the closest to this point is found: this is the zero-shot learning setting defined in [6].

Previous studies defined this semantic space "by hand". [6] manually designed a 218 dimensional representation space in which a concept representation is defined according to the answers to 218 questions such as 'is it manmade?' or 'can you hold it?'. Such a semantic space was designed for a particular set of concepts. Later on, to extend these methods to deal with a larger number of concepts, researchers tried to leverage information from lexical and corpus resources to automatically design a universal and accurate semantic space. For instance, [6] built a 5000 dimensional semantic space from the Google n-gram corpus, [7] found that co-occurrences counts with very high frequency words were an informative representation of words for semantic tasks, [8] examined various semantic feature representations of concrete nouns derived from 50 million English-language webpages, etc.

This work deals with the problem of automatically designing a semantic space for [6]'s task, i.e. predicting the concept from the fMRI image in the zero-shot learning setting. Since previous studies have shown the superiority of manually designed semantic spaces we propose to combine multiple and diverse semantic spaces, either automatically learned from huge corpora, following recent works in the machine learning and representation learning community [9], or designed from various linguistic resources (e.g. *WordNet* [10]). In order to exploit these semantic spaces efficiently, we propose to use an effective blockwise regularized learning algorithm [11] that prevents overfitting and focuses on relevant information contained in the fMRI images.

2 Learning Models for Brain Decoding

Our idea consists in combining multiple semantic spaces, some of them being designed automatically using linguistic resources while others are learned using representation learning ideas such as the one in [9]. Our system for inferring a

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8. Available from http://www.i6doc.com/en/.

concept from an fMRI image is illustrated in figure 2. It relies on two multilinear mapping functions: Ω maps a single word w in a continuous p-dimensional space so that $\Omega(w) = \mathbf{z} \in \mathbb{R}^p$. We refer to this space as a semantic space. Ω is built using external resources [9, 12] and the representation spaces considered are detailed in section 3. β enables us to make the link between the fMRI image vector $\mathbf{x} \in \mathbb{R}^d$ (an image made of d voxels) and the word semantic representation $\mathbf{z} \in \mathbb{R}^p$. We first consider a simple strategy by learning independently multiple ridge regressions: this will be our baseline when considering multiple semantic spaces in our experiments. We then investigate a more advanced multitask strategy using the multitask blockwise regularized LASSO from [11]: by regularizing jointly all regression models, we can take into account globally the relevance of every voxel with respect to the task. This strategy is explained here. Note that one independent model is learned for each subject.

Let $X = \{\mathbf{x}_i\}_{i=1,\dots,N}$, $\mathbf{x}_i \in \mathbb{R}^d$ be the collection of fMRI images for a subject and $Z = \{\mathbf{z}_i\}_{i=1,\dots,N}$, $\mathbf{z}_i \in \mathbb{R}^p$ the collection of associated word semantic representations. The Ridge Regression (RR) consists in learning $\beta \in \mathbb{R}^{d \times p}$ coefficients that map efficiently from the voxel space to the semantic space. For the Multitask LASSO (MTL), the global blockwise regularized problem is formulated as:

$$\underset{\beta}{\operatorname{argmin}} \left(\frac{1}{2} \sum_{i=1}^{N} \|\mathbf{z}_i - \mathbf{x}_i \beta\|^2 + \lambda \sum_{j=1}^{p} \|\beta_j\|_\infty \right) \text{ with } \|\beta_j\|_\infty = \max_{\ell} |\beta_{\ell j}| \quad (1)$$

During training, entire rows of the resulting β matrix will "vanish" so as to focus only on relevant voxels.

After training K models $\beta^{(k)}$, corresponding to K different semantic spaces, we still have to build a decision criterion to choose the word to be associated to the fMRI. Each word w is mapped in the kth semantic space using the $\Omega^{(k)}$ function, thus we get $\Omega^{(k)}(w) \in \mathbb{R}^p$. In parallel, we obtain K semantic representations associated to the fMRI images $\mathbf{x}$ using $\beta^{(k)}$ coefficients. Their cosine similarity can then be computed in the intermediate space and the results are merged using a linear combination, as follows:

$$sim(\mathbf{x}, w) = \sum_{k=1}^{K} \lambda_k \frac{\langle \mathbf{x}\beta^{(k)}, \Omega^{(k)}(w) \rangle}{\|\mathbf{x}\beta^{(k)}\| \; \|\Omega^{(k)}(w)\|} \text{ s.t. } \sum_k \lambda_k = 1 \quad (2)$$

Obviously, the word with the highest similarity to an fMRI image is chosen.

3 Experiments and Discussion

3.1 fMRI Dataset and Task

The fMRI data was collected from nine participants while subjected to a pair of stimuli: a line-drawing depicting a particular concept alongside a text label [5][1] .

[1] [5]'s data is publicly available at http://www.cs.cmu.edu/afs/cs/project/theo-73/www/science2008/data.html. [6, 8, 11] also test their models on that same dataset

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.

There were 60 concepts (classes) belonging to 12 semantic categories (mammals, body parts, buildings, furniture, etc.), each presented 5 times to the participants. Every fMRI image is comprised of about $20,000$ voxels representing the cortex activity. Following [5] we considered in our experiments subsets of 500 to 10000 voxels using the same selection procedure they did, based on a stability criterion.

We investigated the zero-shot learning setting defined in [6]. Models' evaluation was done in a leave-2-out cross-validation setting: the training data consisted in 58 classes, and the models learned were tested on the remaining 2 (thus, the classes of the test set are completely unknown to the models).

3.2 Word Semantic Features

We now describe the three considered approaches to design a semantic space.

WordNet based semantic space (WN) WordNet provides easy ways for designing a semantic space. This lexicon is organized as a hierarchical tree of concepts and subconcepts. As a consequence, it is possible to compute a path in the tree between two concepts (words). Intuitively the smaller this path, the closer these two concepts are [10]. Based on such a metric, a given word can be represented in a fixed p-dimensional space by computing its distance to a given set of p representative words: we considered the most common words in Wikipedia. We will call such a semantic space WN_{path}. Alternative metrics have been proposed in the literature that lead to other semantic spaces: the closeness of two concepts can be measured in respect to their closest common ancestor [12] (let us note the corresponding semantic space WN_{anc}) and [13] defines a criterion inspired from mutual information, comparing the weights of subtrees associated to each concept (this semantic space will be denoted WN_{mi}).

Word2Vec semantic space (W2V) Representation learning has emerged in the recent years as a key research field in the machine learning community. Word2Vec is an efficient tool that learns continuous and dense representations of words from text data [9]. It is a supervised learning approach based on neural networks in which a hidden layer is used to encode a vector representation that captures syntactic and semantic patterns of words.

Human218 semantic space (H_{218}) The last semantic space we considered is a baseline noted H_{218}. As explained before it is a manually designed space which has been obtained from crowdsourcing [6]. For each concept considered a 218 dimensional representation is defined according to the answers from a set of volunteers to 218 questions like *is it manmade?* or *can you hold it?*.

3.3 Results and Discussion

In a preliminary experiment, the semantic space dimension p was optimized using a large set of voxels (we fixed $d = 2000$). Results are reported in Fig 2 : a dimension $p = 150$ seems to offer a good trade-off between complexity and

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.

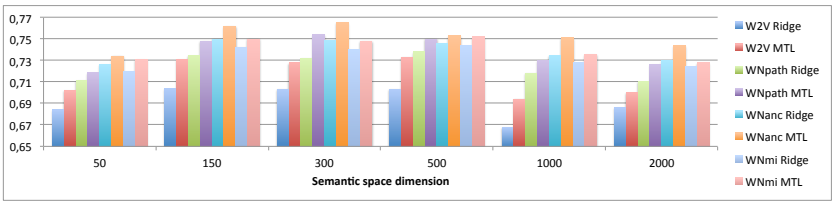

Fig. 2: Accuracy (zero-shot learning) wrt the semantic space dimension, for various semantic spaces (W2V, WN, ...) and for the two training strategies (MTL = Multitask LASSO/ Ridge = Ridge Regression)

accuracy, and this value was kept for further experiments. Also, in every semantic space configuration, we notice that MTL (multitask LASSO) systematically overcomes Ridge regression which validates our regularization strategy for identifying and neglecting unnecessary voxels. Hence we will focus on this model in further experiments.

We then performed a combined experiment to study the impact of voxel preprocessing (reducing the voxel space using the stability criterion proposed in [5]) as well as the interest of mixing different semantic spaces. All results are provided in Fig 3. Best results are obtained for a voxel space size of 2000: we can see MTL procedure can't deal efficiently with large dimensional noisy data such as fMRI images (with their 20000 voxels), and that such a preprocessing to select relevant voxels is recquired.

Our most important result lies in the overall performance on this difficult brain-reading task: up to now, state-of-the-art results relied on Human218 (H_{218}) resources [6], which is hand-made for this task and questions the ability to generalize the process to a larger vocabulary. We demonstrate here the interest of combining different lexical and learned resources to outperform this strategy. While H_{218} reaches an accuracy of 80.3% (last column of Fig. 3), being far above the best single model ($WNanc$) that reaches 76.2%, it is outperformed by our combination schemes. Combining 2 resources provides a significative improvement to catch up with H_{218}: $W2V + WNanc$ model reaches 80.3% accuracy. Adding a third resource ($WNpath$), we reach 80.7% accuracy.

The comparison of various combinations confirms our assumption: it is more relevant to combine heterogeneous spaces like $W2V$ and WN than to work with a single resource.

4 Conclusion

Predicting a concept stimulus from an fMRI image is a hard task which is traditionally tackled through defining a manual semantic space and learning a regression model. While this approach has proved effective for a limited set of concepts, the manual design of the semantic space prevents the approach to be extended to a larger number of concepts. We tackled the problem by relying on multiple

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.

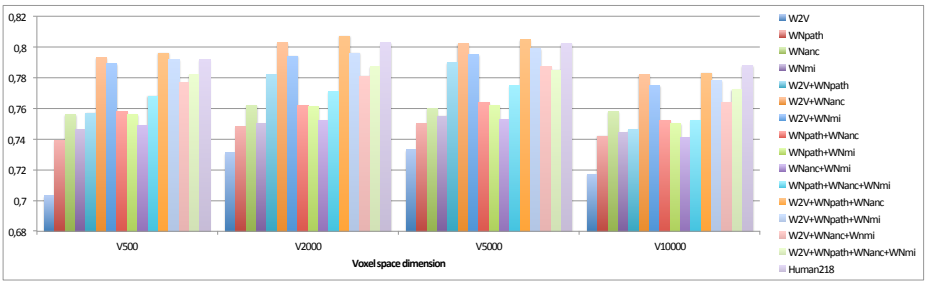

Fig. 3: Accuracy (zero-shot learning) with Multitask LASSO wrt the voxel space dimension and for various semantic space combinations.

semantic spaces automatically designed from resources and trained from large corpora. Given the dimension of fMRI images, it is necessary to implement a robust learning strategy: the multitask LASSO we designed allows us to efficiently select relevant voxels. MTL, combined with Word2Vec and WordNet, catches up with the state-of-the-art performance in brain-reading relying on hand-made resource. It is a promising step towards more advanced brain-reading tasks.

References

[1] R.A. Poldrack. The role of fmri in cognitive neuroscience: where do we stand? *Current opinion in neurobiology*, 18(2):223–227, 2008.

[2] F. Pereira and M. Botvinick. A systematic approach to extracting semantic information from functional mri data. In *NIPS*, pages 2267–2275, 2012.

[3] K.N. Kay, T. Naselaris, R.J. Prenger, and J.L. Gallant. Identifying natural images from human brain activity. *Nature*, 452(7185):352–355, 2008.

[4] D.R. Hardoon, J. Mourao-Miranda, M. Brammer, and J. Shawe-Taylor. Unsupervised analysis of fmri data using kernel canonical correlation. *NeuroImage*, 37(4), 2007.

[5] T.M. Mitchell, S.V. Shinkareva, A. Carlson, K.M. Chang, V.L. Malave, R.A. Mason, and M.A. Just. Predicting human brain activity associated with the meanings of nouns. *science*, 320(5880):1191–1195, 2008.

[6] M. Palatucci, D. Pomerleau, G.E. Hinton, and T.M. Mitchell. Zero-shot learning with semantic output codes. In *NIPS*, 2009.

[7] J.A. Bullinaria and J.P Levy. Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior research methods*, 39(3), 2007.

[8] B. Murphy, P. Talukdar, and T. Mitchell. Selecting corpus-semantic models for neurolinguistic decoding. In *ACL JC on Lexical and Computational Semantics*, 2012.

[9] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv*, 2013.

[10] C. Leacock and M. Chodorow. Combining local context and wordnet similarity for word sense identification. *WordNet: An electronic lexical database*, 49(2), 1998.

[11] H. Liu, M. Palatucci, and J. Zhang. Blockwise coordinate descent procedures for the multi-task lasso, with applications to neural semantic basis discovery. In *ICML*, 2009.

[12] P. Resnik. Using information content to evaluate semantic similarity in a taxonomy. *IJCAI*, 1995.

[13] Z. Wu and M. Palmer. Verbs semantics and lexical selection. In *ACL*. ACL, 1994.