IR Principles for Content-based Indexing and Retrieval of Functional Brain Images

Bing Bai, Paul Kantor
Rutgers University
bbai, kantorp@cs.rutgers.edu

Nicu Cornea, Deborah Silver
Rutgers University
cornea, silver@ece.rutgers.edu

ABSTRACT

In this paper, we explore the concept of a "library of brain images", which implies not only a repository of brain images, but also efficient search and retrieval mechanisms that are based on models derived from IR practice. As a preliminary study, we have worked with a collection of functional MRI brain images assembled in the study of several distinct cognitive tasks. We adapt several classical IR methods (inverted indexing, TFIDF and Latent Semantic Indexing(LSI)) to content-based brain image retrieval. Our results show that efficient and accurate retrieval of brain images is possible, and that representations motivated by the IR perspective are somewhat *more* effective than are methods based on retaining the full image information.

Keywords

fMRI, functional brain image, content-based retrieval, inverted files, TFIDF, Latent Semantic Indexing

1. INTRODUCTION

fMRI (functional Magnetic Resonance Imaging) [3] is a technique used to "monitor" brain activities. In this technique, the intensity of an image element (it is called a *voxel* in 3D images) is related to the level of blood oxygen in the corresponding region. When a cognitive process involves a specific brain region, more oxygen is brought in by blood flow soon after, and this change will brighten the brain region in the image. In a typical fMRI experiment, the experimental subject is assigned a certain type of cognitive task at some scheduled moments, A 3D brain scan is made every few seconds, during each experimental session. By comparing the intensities of voxels "during task" and "being idle", we can estimate which brain regions are "activated" by the cognitive process.

As the size of the (world) data corpus increases, efficient data *sharing* schemes for fMRI data become more and more desirable. Data sharing, of course, provides a larger database for testing and validating analytical algorithms. More importantly, different researchers may find different value in the same data, discovering similarities in the brain's activity, when the cognitive tasks do not seem to be related, based on psychological reasoning alone.

In this paper, we report some investigations of *content-based* indexing of fMRI images, which is validated by using *condition* labels. In other words, when a "query" fMRI image is presented, we ask whether we can return images that represent the same or similar cognitive processes. The potential applications include, but are not limited to, the following: 1) Helping researchers find similar studies and related research work. 2) Helping researchers discover hidden similarities among superficially different studies. 3) Helping doctors diagnose brain disorders, by looking at the clinical history of persons with similar fMRI patterns.

With this correspondence in mind, we show that using the classical information retrieval techniques of inverted indexing (applied not to terms, but to "activated voxels"), retrieval by condition can be implemented effectively and efficiently. The inverted index provides the infrastructure for many different Information Retrieval algorithms. In this note we examine two of the most basic, TFIDF similarity [7] and LSI representation and similarity [2].

2. BACKGROUND

The most widely used mathematical method for analysis of fMRI is the general linear model (GLM)[3]. In the GLM, the time series of intensity values at each voxel is modeled as a linear combination of explanatory variables and Gaussian noise. Each explanatory variable is generated by convolving the stimulus time series with a specific impulse response function, referred to as hemodynamic response function (HRF). The mean and variance of the weight (in the regression) of each explanatory variable is calculated using a linear regression. By comparing each weight with its estimated variance one can estimate the (Fisher) t-statistic, which will be referred to as the t-value of the weight. This is a measure of the degree to which the activation level is influenced by the corresponding stimulus.

3. FEATURE SELECTION

We propose a two-stage feature selection method. The first stage is to build "activation maps", in which the value of each voxel indicates the level of activation of the voxel. We currently use t-maps generated by the GLM module of FSL. The second stage selects the "most important" voxels. There is no consensus on what t-value should be considered "large enough", so here we simply take that 1% of the voxels with largest t-values.

Our retrieval model closely parallels classical text retrieval,

Table 1: **Experiments**

Experiment	Size
Oddball: Recognition of an out of place image or sound	8
Event perception: Watching either a cartoon movie or real film of a human being [8]	53
Morality: Making decisions about problem situations having [4] or lacking combinations of moral and emotional content	150
Recall: Study and recall or recognition of faces, objects and [6] locations	189
Romantic: People in love see pictures of their important [1] others, or of non-significant people	30

with these correspondences:

Brain images	*correspond to*	Documents
Activated voxels	*correspond to*	Terms
Query voxels	*correspond to*	Query Terms

With this similarity in mind, we build both forward and inverted indices. In the forward index, each entry is an image, with a link to its activated voxels. In the inverted index, each entry is a unique voxel (in the standard brain), pointing to a list of brain images in which that voxel is activated.

4. SIMILARITY MEASURES

We investigate three methods: 1) *Overlap*: the datasets are ranked according to the overlap of their activated voxels with those in the query image. 2) *Latent Semantic Indexing* (LSI): LSI [2] is a technique to automatically discover similarities between terms and documents, and to apply this information to information retrieval. As noted, 1 percent of the voxels, those with largest t-values are selected to be "activated". If a voxel is activated, the corresponding value in the term/voxel vector is set to 1, otherwise it is set to 0. The standard routine of LSI is applied after that. 3) *TFIDF*: Under TFIDF [7] the weight of a term is the product of two parts: TF(term frequency) and IDF (inverse document frequency). In our scenario, we set TF to be the t-value, while DF is natually translated as the number of images in which a voxel is activated.

5. TESTING SCHEME

In this preliminary research, we use data from 5 different experiments. The list of experiments and brief descriptions are shown in Table 1.

Our testing scheme is built on an information retrieval framework. We use every image as query, and evaluate the performance of our method by checking the returned ranked lists. A retrieved image is considered "relevant" to the query only if they are both for the same experimental condition. As noted, different experiments have different numbers of datasets. In this case, average precision will not behave as one might like, and we are suspicious of using it. Instead,we use the "area under the ROC" [5] (for the sake of simplicity, we call this "ROC area") to evaluate a retrieval method.

6. RESULTS

In Table 2 we list the average ROC area for 4 different similarity measures. We regard "cosine" as the baseline system. "cosine" takes whole t-maps as big vectors where each voxel is an element in these vectors; the similarity between two t-maps is the cosine of the angle between the vectors.

Table 2: **Average ROC area for 430 datasets.**

Method	Mean ROC area	Stdev	Std Error of the Mean
Cosine	.704	.148	.007
Overlap	.733	.152	.007
TFIDF	.735	.149	.007
LSI(10 components)	**.755**	.134	.006

Except for the baseline system, all methods select only 1 percent of the voxels to be labeled activated. Using the Bonferroni correction for multiple comparisons, almost all of the pair-wise differences in Table 2 are significant at the 95% confidence level. The only exception is the pair TFIDF and Overlap. This lack of difference suggests that perhaps moving from binarized voxels to weighted voxels will not be an important factor in retrieval.

We see that all methods based on thresholded "t-maps" perform better than the baseline method significantly. That provides additional support for the essentially arbitrary choice of thresholding by 1%. As the threshold is lowered, we are more likely to be including random noise, which will hurt the quality of retrieval, rather than improving it. More importantly, this thresholding approach makes inverted indexing quite efficient.

The performance of LSI is known to depend on the number of components retained in indexing. In this study we find that the average ROC area reaches its maximum when we consider only the first 10 components.

7. CONCLUSION

In this preliminary study, we show that efficient fMRI retrieval can be implemented by adapting classical textual retrieval techniques and using inverted indexing. The results show that retrieval by condition with high precision is possible.

8. ACKNOWLEDGMENTS

This work is supported by NSF grant ITR 0118760. The authors are grateful to Ali Shokoufandeh and Sven Dickinson for their helpful advice and encouragement.

9. REFERENCES

[1] A. Aron, H. Fisher, D. Mashek, G. Strong, H. Li, and L. Brown. Reward, motivation, and emotion systems associated with early-stage intense romantic love. *J Neurophysiol*, 94:327–337, 2005.

[2] S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman. Indexing by latent semantic analysis. *JASIST*, 41(6):391–407, 1990.

[3] R. Frackowiak, K. Friston, C. Frith, R. Dolan, C. Price, S. Zeki, J. Ashburner, and W. Penny. *Human Brain Function (2nd Edition)*. Elsevier Academic Press, 2004.

[4] J. Greene, R. Sommerville, L. Nystrom, J. Darley, and J. Cohen. An fmri investigation of emotional engagement in moral judgment. *Science*, 293, 2001.

[5] S. Mason and N. Graham. Areas beneath the relative operating characteristics (roc) and relative operating levels (rol) curves: Statistical significance and interpretation. *Q. J. R. Meteorol. Soc*, 30:291–303, 1982.

[6] S. Polyn, J. Cohen, and K. Norman. Detecting distributed patterns in an fmri study of free recall. In *Soc. for Neurosci. conf.*, San Diego, CA, 2004.

[7] G. Salton and C. Buckley. Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, 24(5), 1988.

[8] A. Zaimi, C. Hanson, and S. Hanson. Event perception of schema-rich and schema-poor video sequences during fmri scanning: Top down versus bottom up processing. In *In Proc. Ann. Mtng. Cog. Neuro. Soc.*, 2004.