Methods of Predicting the Brain Activity Based on Noun

Jianhua Jiang
School of Computer and Information
Science, Southwest University
Chongqing 400715, China
email:jjhluxun@163.com

Xu Yu
School of Computer and Information
Science, Southwest University
Chongqing 400715, China
email: dingo@swu.edu.cn

Zhixing Huang
School of Computer and Information
Science, Southwest University
Chongqing 400715, China
email: huangzx@swu.edu.cn

Abstract—Over the last decade, functional magnetic resonance imaging (fMRI) has become a primary tool to predict the brain activity. During the past research, researchers transfer the focus from the picture to the word. The results of these researches are relatively successful. In this paper, several typical methods which are machine learning methods are introduced. And most of the methods are by using fMRI data associated with word's features. The semantic features (properties or factors) support word's neural representation, and have a certain commonality in the people. The purpose of the application of these methods is used for prediction or classification.

Keywords-fMRI; semantic features; machine learning; classification; prediction

I. INTRODUCTION

Over the last decade, functional magnetic resonance imaging (fMRI) has become a primary tool to identify the brain activity. Many of researches have been done based on fMRI image or data for prediction or classification. And the results of these researches are relatively successful. The research area of focus moves from picture to noun (Mitchell et al., 2008; Chang et al., 2010; Just et al.,2010; Gerven et al.,2010; Gershman et al.,2011; Murphy et al.,2012; etc.). In this paper, several typical methods are introduced. These methods are basically by using fMRI data associated with word's features. The semantic features support word's neural representation, and have a certain commonality in the people.

The fMRI data acquisition data and signal processing methods are previously reported in another publication (Mitchell et al., 2008). The procedure of noun (as shown
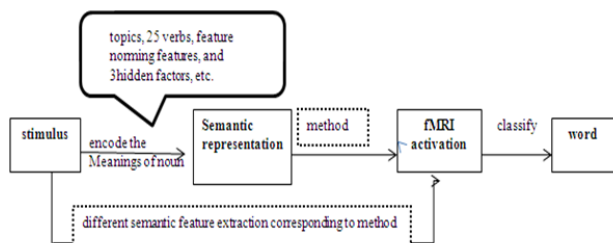

Fig. 1   The procedure of brain activity prediction.

below Fig.1) is classification or prediction is first training on part of the fMRI data set and then testing on the remaining data set. These methods contribute to learn the mapping between concepts based on semantic features and the neural activity (using fMRI data) that get from neuroimaging experiments. Then try to correctly match the predicted (fMRI activation estimated by generated model) and observed activations (observed in the neuroimaging experiment).

The Fig.1 indicates what to use as semantic features, how to extract their values, what and how to predict. First, encode the meaning of the stimulus word in terms of semantic features for semantic representation or neural representation. Then, use the method to learn the mapping between the semantic representation and fMRI activation. In this learning, it includes training and testing. Finally, classify the predicted fMRI activation to word.

II. METHODS

In this paper, several methods (as shown below Fig. 2) are summarized associated with semantic features and machine learning. In these methods, different semantic

Methods of predicting the brain activity based on nouns	References
computational model associated with 25 sensory-motor verbs as semantic features	Mitchell et al.,2008[2]
regression model associated with object features derived from behavioral feature norming	Chang et al.,2010[3]
(GNB)-pooled variance classifier function and linear regression associated with 4 factors derived from two-level factor analysis	Just et al.,2010[4]
Bayesian logistic regression directly associated with word's properties	Gerven et al.,2010[5]
a regularised linear combination associated with semantic features which are extracted in word-region model and word-collocate models in which co-occurrence frequency is normalized by PPMI	Murphy et al.,2012[6]
topographic latent source analysis (TLSA)	Gershman et al.,2011[7]

Fig. 2   Methods of brain activity prediction.

feature extraction is corresponding to its method.

A. Multiple Linear Regression

The first method is computational model associated with 25 sensory-motor verbs as intermediate semantic features. The authors design a predictive model for predicting fMRI activation for arbitrary noun stimuli. In the model, first encode the meaning of the stimuli word in term of semantic features which are the word's co-occurrence frequencies with each of 25 sensory-motor verbs. Then use a trainable computational model to train and test. Finally, get the predicted activity and match it and observed activity. The

usable method is

$$y_v = \sum_{i=1}^{n} c_{vi} f_i(w).$$

(1)

The model mainly is used to predict the fMRI activation y_v at voxel v. In the method, $f_i(w)$ is the value of the ith semantic feature for word w, n is the number of semantic features in the model, and c_{vi} is a learned scalar parameter that specifies the degree to which the ith intermediate semantic feature activates voxel v.

It is a natural progression that the study moves from pretheoretical cataloging of data toward development of computational models and the beginnings of a theory of neural representations [2]. It starts to focus on word rather than only picture when study people's thoughts. Although, the data set (stimulus word and semantic feature) is limited to special area and the authors don't explain how patterns of neural activity encode information in an intermediate semantic representation.

B. *Multivariate Multiple Linear Regression*

On the basis of the method 1, the authors (Chang et al., 2010) propose a generative classifier that models the hidden factors that underpin the neural representation of objects, using a multivariate multiple linear regression model [3]. In the model, the features of word which are retrieved from Cree and McRae (2003) are encoded into BR or DT knowledge types. Then use the method to train (learn the mapping between semantic representation and the fMRI activation) and test (predict the fMRI activation). The method is

$$a_v = \sum_{i=1}^{n} \beta_{vi} f_i(w) + \varepsilon_v.$$

(2)

It is improved based on method 1(Mitchell et al., 2008). ε_v is the model's error term that represents the unexplained variation in the response variable.

Here, the authors design another step (nearest neighbor and SVM classifier) to decode the word associated with observed fMRI activation. In this experiment, the method can generalize to make predictions for a previously unseen word, given the semantic features for that word. The authors prove that features in the semantic representation are more likely to be shared across words. Although, the feature norming in accounting for participants' thoughts when they think about an object is that participants may fail to retrieve a characteristic but psychologically unavailable feature of an object [3].

a)   C. (GNB)-pooled Variance Classifier

Different from method 1, 2, this method is machine learning modeling approaches including GNB and linear regression. The used (GNB)-pooled variance classifier is a mapping function *f* :

$$voxel\ activation\ levels \rightarrow Y_i,\ i=1, …, m.$$

(3)

where Y_i are the 60 words, and *voxel activation levels* are the mean activation levels of the selected voxels. The classification rule is

$$Y \xleftarrow[y_i]{} \arg\max P(Y = y_i) \prod_j P(X_j \mid Y = y_i).$$

(4)

Here the machine learning methods can be implemented by three stages: feature selection (voxels are selected based on semantic factors which are derived from two-level factor analysis method), classifier training (the classifier is (3) which obeys the rule (4)), and testing on within and across participants.

The authors indicate that not only do people have concepts in common, but also their brain coding of the concepts is similar. However, that the location and activation levels did not have to be common across people. The research establishes a new way of describing brain activity (using factor analysis and ML-generative model (GNB and linear regression) to fMRI data for prediction). The stimuli exclusively focus on the representation of words rather than on pictures of objects. It proves the commonality of the neural representation of words across different people. Although, the semantic features are highly subjective and phasic.

D. *Bayesian Logistic Regression*

The above three methods' feature selection (selecting voxel based on semantic features of stimulus word for experiment design, resulting fMRI activation) is in view of using voxels. Rather than analyzing the raw voxel information present in the dataset, the method 4 chose to use the average activation of 116 regions of interest as defined by the Automated Anatomical Labeling (AAL) atlas. ROI activations are used as input to the classifier, facilitating interpretation since results can be analyzed at the ROI level instead of at the single voxel level.

The method 4 uses Bayesian logistic regression as the underlying classification model. The predictive density is

$$p(y \mid x, D, \Theta) = \int B(y; l^{-1}(x^T \beta)) p(\beta \mid D, \Theta) d\beta.$$

(5)

In this experiment, the authors directly use the word's properties to measure BOLD response x, and in this process, get the regression coefficients β. At the same time, observe the training data D=(y, X) and hyper-parameters Θ. Then predict the semantic category y of word from the measured BOLD response x. The classification model is trained using all of the nouns in two of the categories except one. Then predict the semantic category of each of the six images of the excluded noun.

The method achieves good classification performance for many of the category pairs. However, using the voxel data directly would have resulted in better accuracy. Bayesian multi-task learning allows generalization across subjects, and leads to more interpretable models, thereby facilitating the interpretation of the models obtained using multivariate analysis. The most significant contribution is the ability obtain generalized group results using coupled data from all nine subjects. Each subject's model become strongly or at times perfectly correlated with the other models when using strong coupling strengths.

E. *Regularised Linear Combination*

The four methods ahead using semantic features require extensive manual intervention to get. The method 5 initiates a systematic comparison of automatically-derived semantic features representing individual concept. Then train the model with the semantic features, using a regularised linear combination associated with semantic features which are extracted in word-region model and word-collocate models in which co-occurrence frequency is normalized by PPMI. The linear combination is

$$f = C\beta + \lambda \| \beta \|^2 . \tag{6}$$

where f is the vector of activations of a specific fMRI feature for different concepts, the matrix C contains the values of the semantic features for the same concepts, β is the vector of weights we must learn for each of those (corpus-derived) features, and λ tunes the degree of regularisation.[6].

The model can predict the fMRI activation for words that not in the training set. The authors find that dependency parse-based features are the most effective, and the classification performance is better than any published for a corpus-based model. The simple word semantic features with directional information provide a suboptimal solution at much lower computational cost.

F. TLSA (Topographic Latent Source Analysis)

In addition to the above voxel-based models, there is another method: TLSA. The main difference between the first six methods and TLSA is that the hypotheses in TLSA concern latent sources rather than voxels. These latent sources are defined in terms of basis functions over continuous space which concentrates their energy in local regions. The model assumes that fMRI images are generated by a covariate-dependent superposition of latent sources [7]. So the authors give the model a definition to predict the neural activity at each observation:

$$y_{nv} = \sum_c x_{nc} \sum_k w_{ck} f_{kv} + \varepsilon_{nv}. \tag{7}$$

Where C is the number of covariates, N is the number of observations, K is the number of latent sources, and V is the number of voxels. x_{nc} is design matrix containing each covariate's time series. w_{ck} is real-valued weight matrix encoding how each covariate loads on each source. f_{kv} is non-negative real-valued basis image matrix encoding the canonical spatial pattern (over voxels) associated with each latent source. y_{nv} is real-valued fMRI data matrix, the pattern of activity at each observation. $\varepsilon_{nv} \sim N(0, \tau^{-1})$. [7]

At last, the authors evaluate TLSA according to prediction, reconstruction and reproducibility. The performance is as good as other methods. The advantage using latent sources rather than that of voxel as a parameter is that latent sources have intrinsic spatial extent and are suitable for capturing spatial patterns without the need for post-hoc corrections like cluster-size thresholds. However, there is limitation that the authors consider that the number of latent sources to be known and affirmatory.

G. The Results of Methods

For above several methods based on word's semantic features or properties which derive from different way, the classification or prediction performance is relatively good, greater than chance levels. The method proves that semantic features of word have relationship with neural activation associated with the thinking word (Mitchell et al., 2008). The features for learning share certain commonalities across individuals and words ([2], [3], [4], [5], etc.). The generative models can make prediction for a previously unseen word, given the semantic features for that word ([3], [4], [5], etc.). Classification performance is better when comparison between different semantic categories ([4], etc.).

In fact, several methods described in this paper are statistical models (e.g. Bayesian [4], [5]) and linear models (e.g. regression [2], [3], [7], [6]) in view of machine learning techniques. Whether these methods are based on what kind of semantic features or word's meaning, or rely on voxel-based model or not, the prediction accuracy is much greater than chance level (p>0.5). During these methods, the prediction performance of the method 5 is better than others. But, every method has its emphasis and brings innovation.

III. SUMMARY

In this paper, the introduced methods are using fMRI data, because fMRI data offers direct insight into how the brain is functioning in response to given stimuli. The dataset except from fMRI data, there exists research data from EEG or MEG. The methods prove that predicting the brain activity based on the semantic features of nouns using the fMRI data is feasible. The semantic features underlying the neural representation of concrete nouns are relating to brain locations and show the commonality across people. So far, most of research methods belong to the machine learning methods, and get a certain degree of success. When the prediction performance is less than 0.84, many researchers believe that in addition to improvements in the method, but also that the used fMRI image data has a certain noise. During the past research, researchers transfer the focus from the picture to the word. Although, the pictures and words trigger different brain activity, they also have commonality in neural representations.

REFERENCES

[1] Kendrick N Kay, Jack L Gallant, "I can see what you see", nature neuroscience, 12(3), 2009.

[2] Tom M. Mitchell, Svetlana V. Shinkareva, Andrew Carlson, Kai-Min Chang , Vicente L. Malave, "Predicting human brain activity associated with the meanings of nouns", Department of Psychology, Paper (225), 2008.

[3]Kai-min Kevin Chang, Tom Mitchell, Marcel Adam Just, "Quantitative modeling of the neural representation of objects: How semantic feature norms can account for fMRI activation", NeuroImage, Paper(56), 716–727, 2011.

[4] Marcel Adam Just, Vladimir L. Cherkassky, S Aryal, Tom M. Mitchell, "A Neurosemantic Theory of Concrete Noun Representation Based on the Underlying Brain Codes", Department of Psychology, Paper (319), 2010.

[5] Marcel van Gerven, Irina Simanova, "Concept Classification with Bayesian Multi-task Learning", NAACL HLT 2010 First Workshop on Computational Neurolinguistics, 10–17, 2010.

[6] Brian Murphy, Partha Talukdar, Tom Mitchell, "Selecting Corpus-Semantic Models for Neurolinguistic Decoding", First Joint Conference on Lexical and Computational Semantics, 114–123, 2012.

[7] Samuel J. Gershman, David M. Blei, Francisco Pereira, Kenneth A. Norman, "A topographic latent source model for fMRI data", NeuroImage, Paper (57), 89–100, 2011.

[8] Francisco Pereira, Greg Detre, Matthew Botvinick, "Generating descriptive text from functional brain images", Human Neuroscience, 5 (72), 2011.

[9] Francisco Pereira, Tom Mitchell, Matthew Botvinick, "Machine learning classifiers and fMRI: a tutorial overview", NeuroImage , pages S199-S209, 2009.

[10]FranciscoPereira, Matthew Botvinick, GregDetre, "Learning semantic features for fMRI data from definitional text", NAACL HLT 2010 First Workshop on Computational Neurolinguistics, pages 1–9, 2010.

 [11] Colin Kelly, Barry Devereux, Anna Korhonen, "Acquiring Human-like Feature-Based Conceptual Representations from Corpora", NAACL HLT 2010 First Workshop on Computational Neurolinguistics, pages 61–69,2010.

[12] Barry Devereux, Colin Kelly & Anna Korhonen, "Using fMRI activation to conceptual stimuli to evaluate methods for extracting conceptual representations from corpora", NAACL HLT 2010 First Workshop on Computational Neurolinguistics, pages 70–78, 2010.

[13] Fabian M. Suchanek, Gjergji Kasneci, Gerhard Weikum, "YAGO: A Core of Semantic Knowledge Unifying WordNet and Wikipedia", WWW 2007, pages697-706, 2007.

[14]Ahmad Babaeian Jelodar, Mehrdad Alizadeh, Shahram Khadivi, "WordNet Based Features for Predicting Brain   Activity associated with meanings of nouns", NAACL HLT 2010 First Workshop on Computational Neurolinguistics, pages 18–26, 2010.

[15] Svetlana V. Shinkareva, Vicente L. Malave, Marcel Adam Just, Tom M. Mitchell, "Exploring commonalities across participants in the neural representation of objects", Human Brain Mapping 33: 1375-1383, 2012.

[16] Mark Steyvers, "Combining feature norms and text data with topic models", Acta Psychologica 133, 234–243, 2010.

[17] Daniel Cerato Germann, Aline Villavicencio, Maity Siqueira, "An Investigation on Polysemy and Lexical Organization of Verbs", NAACL HLT 2010 First Workshop on Computational Neurolinguistics, pages 52–60, 2010.

[18] David M. Blei, John D. Lafferty, "Correlated Topic Models", Advances in Neural Information Processing Systems, paper (18), 147—154, 2006.

[19]Svetlana V. Shinkareva, Vicente L. Malave, Robert A. Mason, Tom M. Mitchel, Marcel Adam Just, "Commonality of neural representations of words and pictures", NeuroImage, paper (54), 2418–2425, 2011.

[20]Alona Fyshe, Emily Fox, David Dunson, Tom Mitchell, "Hierarchical Latent Dictionaries for Models of Brain Activation", 15th International Conference on Artificial Intelligence and Statistics, Volume XX of JMLR: W&CP XX, 2012.