Multitask machine learning for brain-state classification

Yash S Shah[1], Ashish Farmer[1], Luis Hernandez-Garcia[1], Douglas C Noll[1], Mark Greenwald[2], Jon-Kar Zubieta[1], and Scott J Peltier[1]
[1]*University of Michigan, Ann Arbor, MI, United States,* [2]*Wayne State University, Detroit, MI, United States*

Introduction

In recent times, machine learning techniques have been used increasingly for fMRI data analysis especially for brain-state classification purposes [1]. They allow users to learn patterns from examples of labeled brain volumes and then predict the state of a previously unseen brain volume [2]. Standard single-task SVM has been used for such tasks in recent studies for real-time neurofeedback applications [3]. In this standard formulation, we do not exploit the relatedness in the brain activation of different subjects in response to similar stimuli. Hence, we cast the problem as a multitask machine learning classification problem as described in [4,5]. Each subject's brain activation data is considered as a separate task and it is believed to be related to every other task to some extent which helps in building a generalized classifier. In our study, we present promising results for classifying craving and non-craving brain states of nicotine dependent subjects by taking advantage of the similarity of the related tasks.

Methods

Subject screening: 20 subjects were scanned after overnight abstinence verified by measuring decrease in CO level on day of scan. Several measures including expired CO level (atleast 12ppm), number of cigarettes smoked daily (atleast 10), Fagerström scores (atleast 4) were considered to ensure moderate smoking and level of nicotine dependence.

Data acquisition and Paradigm: Bold functional images were collected on a 3T GE scanner using T2*-weighted single-shot custom spiral-in sequence. (TR/TE/FA/FOV=2s/30ms/90°/22cm, 64x64 matrix, 40 axial slices of 3mm thickness). Paradigm included images that were previously used in [6,7] that depicted smoking and non-smoking scenes and were presented in alternating blocks to induce craving or suppress it. (20s blocks, 5 pictures for 4s each, 16 repeats, with 4s static fixation image in between each block, 384s total time).

Dimensionality reduction: We employ the two-sample t-test score as a univariate feature subset selection technique. All the features are ranked according to their ability to discriminate between the two classes and then only a subset of these features is selected. We define the t-score for a feature 'i' in the data as:

$$t(i) = \frac{\mu A(i) - \mu B(i)}{\sqrt{\frac{\sigma A(i)}{|A|} + \frac{\sigma B(i)}{|B|}}}$$

where, A and B are the two separate classes. μA, σA, |A| and μB, σB, |B| denote the average, variance and number of data points in class A and B respectively.

Only the first k features which have the highest t-scores and thus are most discriminatory are then selected as our data. Optimal value of k was determined to be 1000 using cross validation. Hence the entire data was reduced to have only 1000 features or voxel intensities.

SVM classification: 3dSVM plugin [8] in AFNI [9] was used to perform the SVM classification. A linear kernel was used for the standard single task SVM whereas a separate kernel was defined for the multitask formulation. Here, T is the number of subjects or tasks used for training and M is the number of randomly picked samples or training volumes per task.

Single task kernel : $K\{x,s\} = x^T * s$; where, x and s are two samples of the 1000-dimentional data.

Multitask kernel : $K\{(x, t), (s, q)\} = (1 - \lambda + \lambda * T * \delta_{tq}) * x^T * s$; $1 \le t,q \le T$

data point x from task t is represented as an ordered pair (x,t)
δ_{tq} is 1 when t=q, and 0 otherwise

In single task SVM, we build a separate classifier for each task and then to predict the class of each test volume, we take a majority vote over all these classifiers. In the multitask formulation, by using a multitask kernel function as defined above, learning many related tasks simultaneously can be cast as a single task learning problem and only one common classifier is obtained which is used to predict the class of the test volume. This kernel models relations among the tasks by employing a coupling parameter lambda. When lambda equals 0, each task influences every other task heavily and makes the problem similar to building one common classifier for all tasks together whereas when lambda equals 1, it implies that the tasks are all learnt independently and that an individual SVM classifier is learnt per task. Thus, the choice of lambda is an important decision in multitask problems.

Experimental iterations: Several iterations of the multitask problem were carried out by training on T=5, 10 and 15 tasks (subjects) and randomly picking M=20, 40 and 60 samples (training points) per task and varying lambda over the entire range from 0 to 1. Testing was performed on totally separate previously unseen data.

Results

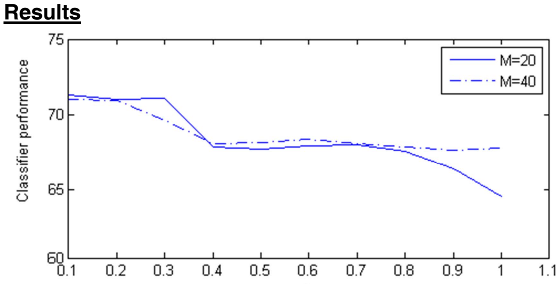

Figure 1: Classifier performance v/s lambda for 20 and 40 randomly picked samples per task

Tasks	Samples per task	Individual SVM	MT-SVM
5	20	56.50	62.50
5	40	70.63	63.25
5	60	72.25	70.38
10	20	57.50	70.56
10	40	66.81	70.38
10	60	69.81	70.81
15	20	58.38	71.08
15	40	65.50	71.00
15	60	70.29	71.17

Table 1: Classifier performance of individual SVM and MT-SVM over different number of tasks (subjects) and samples per task for λ=0.2

Discussion and Conclusion

As it is evident from figure 1, the optimal classifier performance was obtained when the data points were closely coupled with a value of lambda = 0.2 which was used for all further tests. In table 1, we observe that for small number of tasks, proposed multitask SVM gives inferior performance and one might prefer using single task SVM. In accordance with previous studies on multitask learning, multitask classifier performance improves with an increase in the number of tasks used for training the model eventhough data from each task may be highly subsampled by picking only a few of the samples per task. When the model is trained on 10 or 15 subjects, multitask learning performs as good as or better than single task SVM. This finding is noteworthy because in this new formulation, the classifier is trained using data not only from the same subject but other related subjects as well. This helps in building a generalized classifier so that test volumes from newer subjects can be classified without ever training on that subject before. As demonstrated, multitask learning presents an accurate learning of multiple tasks simultaneously by making use of the similarity in the related tasks.

References

[1]Pereira, et al. (2009) Neuroimage, 45:199; [2]Mitchell, et al. (2004) ML, 57:145; [3]LaConte, et al. (2007) HBM, 28: 1033; [4]Caruana, et al. (1997) ML, 28:41; [5]Evgeniou, et al. (2005) JML, 6:615; [6]Gilbert, et al. (2003) Emotional image series v1.1; [7]King, et al. (2010) Neuropsych 35(3):692; [8]LaConte, et al. (2005) Neuroimage, 26:317; [9]Cox (1996) Comp. and Biomed. Res., 29:162