How To Use the Multi-Voxels Pattern Classification Toolbox (MVPC Toolbox)

•General overview

What we can do with this toolbox ?

The aim of this toolbox is to make easy to perform the fMRI pattern classification analysis with multiple algorithms.

You can perform below analyses with this toolbox.

1) Preprocessings including conversion from DICOM files to NIFTI files and usual preprocessings for fMRI data (slice timing correction, co-registration, normalization etc.)
2) Voxel and volume selection for decoding.
3) fMRI decoding by using various algorithms.
4) Leave-one-session-out and leave-two-session-out cross validation for performance evaluation and adjustment of (a) hyper-parameter(s).
5) save the results graph of pattern classification analyses.

If you want to use the script for pattern classification algorithm only, please use `clsfy.m`. For the description of `clsfy.m`, please see “README_clsfy”.

Basic usage

All run-level scripts requires one structure as input, `MY_VAR`, which contains almost all required parameters for analyses. Example of and instruction about the parameters in `MY_VAR` was shown in the files myv_... .m in the Example -> batch_template directory.

Feedback and bug report

Any feedback and bug reports are welcome. Please mail to satoshi.hirose [at] nict.go.jp (please replace the [at] with the ‘@’ symbol).

Copyright

MVPC toolbox is free but copyrighted software, distributed under the terms of the GNU General Public Licence. Further details on copyleft can be found at http://www.gnu.org/copyleft/. No formal support or maintenance is provided or implied.

•Installation

To install MVPC Toolbox, just unzip the file and add all the subdirectories to MATLAB path.
For full use of this toolbox, the open-souse MATLAB toolboxes listed below are required. For the installation of these toolboxes, please see their instructions.

Required toolboxes

Name	Description	Site
SPM 5	The toolbox for analyzing the fMRI, PET, etc. data. In our toolbox, SPM is mainly used for preprocessing and univariate analysis for ROI selection	http://www.fil.ion.ucl.ac.uk/spm/
Volume Toolbox*	The toolbox for dealing with SPM fMRI volume files. Mainly used for extracting ROI data from image files.	http://sourceforge.net/projects/spmtools/
Anatomy Toolbox	This toolbox contains human probabilistic cytoarchitectonic maps derived from vast of studies. Mainly used for defining the ROI with anatomical structure.	http://www.fz-juelich.de/inm/inm-1/DE/Forschung/_docs/SPMAnatomyToolbox/SPMAnatomyToolbox_node.html
LIBSVM	Toolbox for SVM.	http://www.csie.ntu.edu.tw/~cjlin/libsvm/
SLR Toolbox	Toolbox for SLR.	http://www.cns.atr.jp/~oyamashi/SLR_WEB.html
Glmnet in MATLAB	Toolvox for LASSO and Elastic Net.	http://www.stanford.edu/~hastie/glmnet_matlab/

*A directory in Volume Toolbox “Single_Volumes” should be added to the MATLAB path.

•Future Release

So far, the followings are not available (please wait future release).

1) Weight-based functional mapping
2) Beta-values (or other statistical values) based decoding
3) GUI for setting parameters

For other requests please contact us.

•Quickest how to's

1. Prepare fMRI data (DICOM files).
2. Prepare psych directory (see the next page), including parameter files for SPM analysis and parameter files for decoding analysis.
3. Copy and Paste `my_script_all.m` and `myv_all.m` in Example -> Batch_for_all_analyses to somewhere.
4. Edit `myv_all.m` as you like.
5. Run my_script_all.m.
6. Please wait until process completed.

•Learn the usage

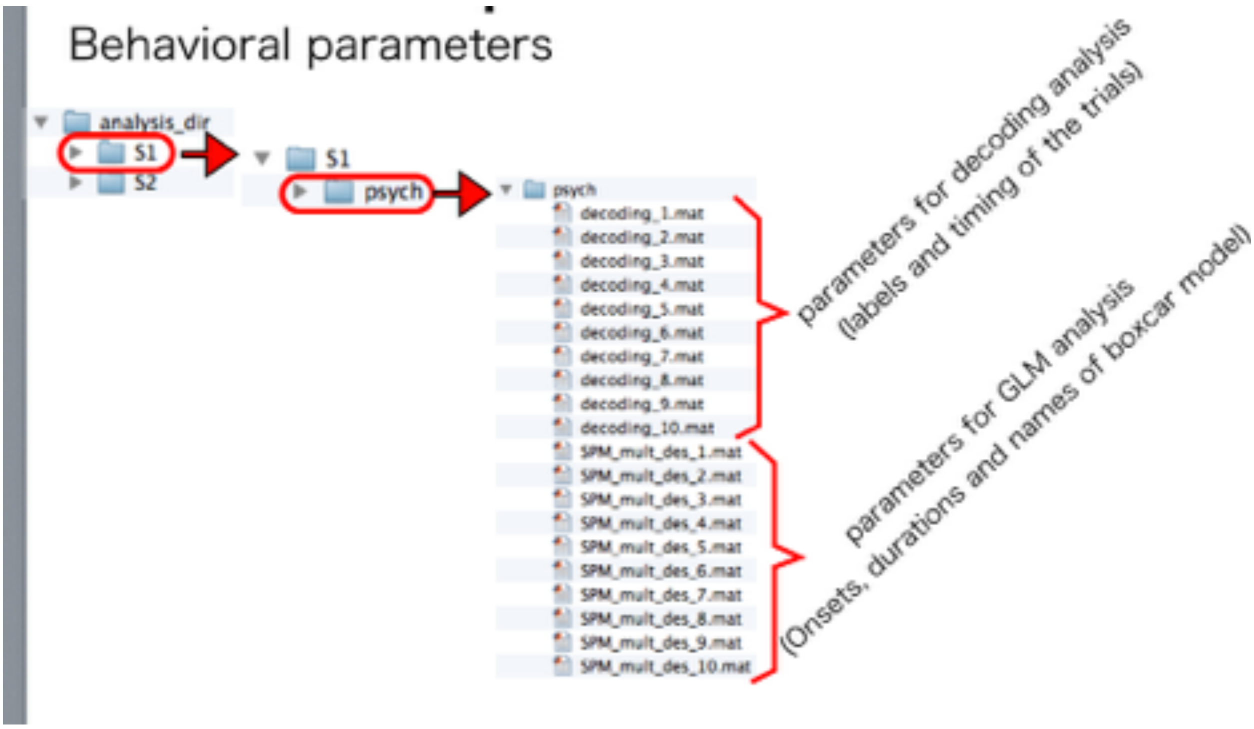
The tutorial may help you to understand what is done in each process.

•Preparation

Our toolbox requires a certain directory structure. Please prepare the appropriate directories and files for the analysis (see below). Behavioral data (information about labels and trials) and SPM models should be saved with the required format. (See Example in Tutorial). Most of the parameters for analyses can be determined in the structure of `MYV` (see `myv_….m` in Tutorial).

- <u>Directory structure</u>

The "psych" directory is required for each participants. Two types of mat files should be included in the directory.



Behavioral parameters

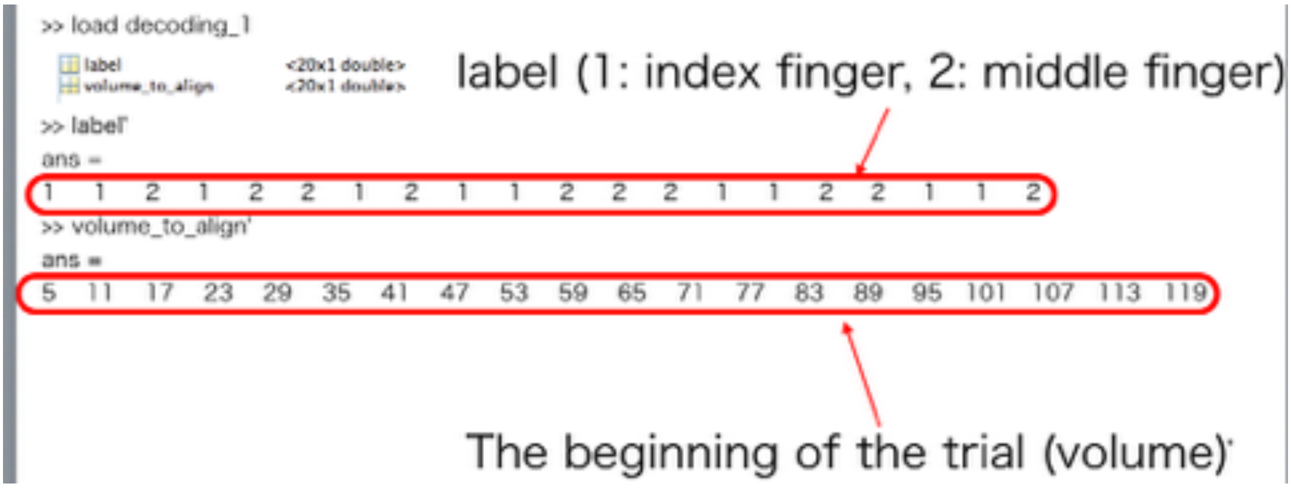
```
analysis_dir
    S1
    S2

S1
    psych

psych
    decoding_1.mat
    decoding_2.mat
    decoding_3.mat
    decoding_4.mat
    decoding_5.mat
    decoding_6.mat
    decoding_7.mat
    decoding_8.mat
    decoding_9.mat
    decoding_10.mat
    SPM_mult_des_1.mat
    SPM_mult_des_2.mat
    SPM_mult_des_3.mat
    SPM_mult_des_4.mat
    SPM_mult_des_5.mat
    SPM_mult_des_6.mat
    SPM_mult_des_7.mat
    SPM_mult_des_8.mat
    SPM_mult_des_9.mat
    SPM_mult_des_10.mat
```

parameters for decoding analysis (labels and timing of the trials)

parameters for GLM analysis (Onsets, durations and names of boxcar model)

- decoding_1.mat ...

The files for pattern classification.
Each files corresponds to each session for decoding analysis.
For detail, please see section of ROI setting.



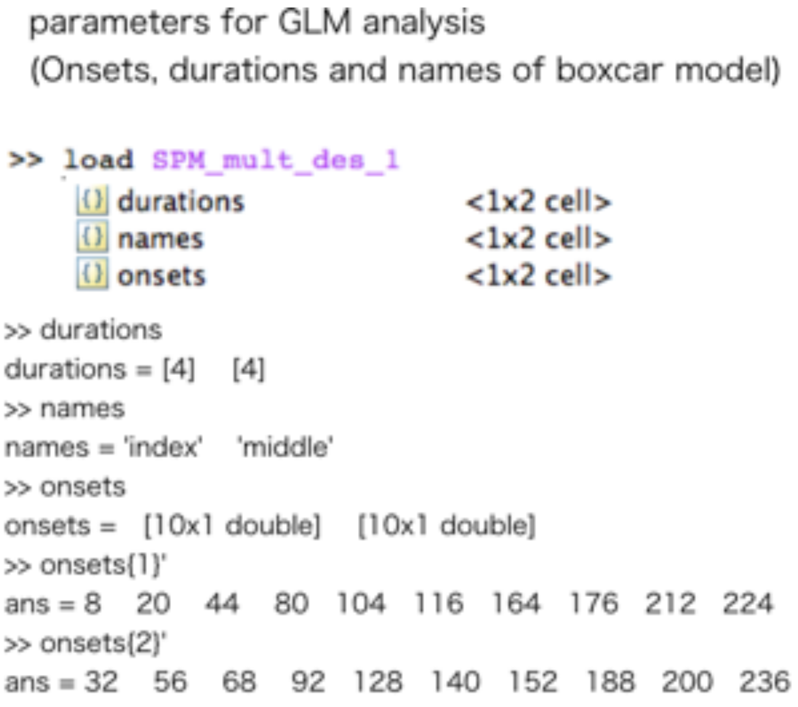
```
>> load decoding_1
  label              <20x1 double>
  volume_to_align    <20x1 double>
>> label'
ans =
 1   1   2   1   2   2   1   2   1   1   2   2   2   1   1   2   2   1   1   2
>> volume_to_align'
ans =
 5  11  17  23  29  35  41  47  53  59  65  71  77  83  89  95  101  107  113  119
```

label (1: index finger, 2: middle finger)

The beginning of the trial (volume)'

- SPM_mult_des_1.mat ...

The files for SPM analysis.
Each files corresponds to each session for SPM analysis.
For detail, please see page 65 of SPM manual http://www.fil.ion.ucl.ac.uk/spm/doc/manual.pdf for detail



parameters for GLM analysis
(Onsets, durations and names of boxcar model)

```
>> load SPM_mult_des_1
    durations             <1x2 cell>
    names                 <1x2 cell>
    onsets                <1x2 cell>
>> durations
durations = [4]    [4]
>> names
names = 'index'    'middle'
>> onsets
onsets =    [10x1 double]    [10x1 double]
>> onsets{1}'
ans = 8    20    44    80    104    116    164    176    212    224
>> onsets{2}'
ans = 32    56    68    92    128    140    152    188    200    236
```

•Preprocessing

- Conversion of fMRI data form DICOM format to NIFTI format

Script

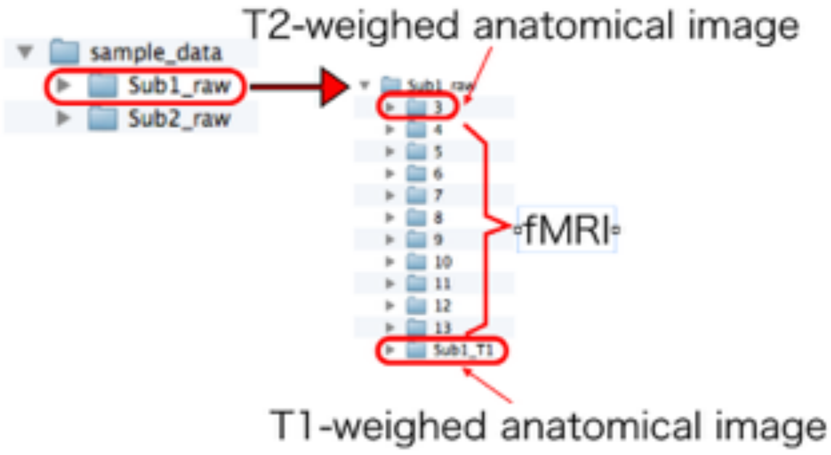
```
convert_DICOM_to_NIFTI(MY_VAR)
```

Example of Batch files

```
myv_convert_DICOM_to_NIFTI.m
my_script_convert_DICOM_to_NIFTI.m
```

Preparation

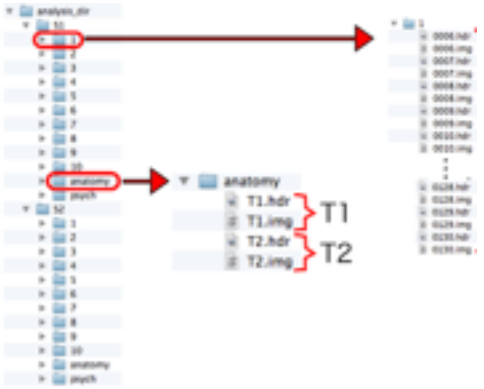
DICOM files for each participant should be in a directory. Sub-directories should be organized as follows (The directory names are arbitrary).



The DICOM files obtained from each fMRI runs should be included in a sub-directory. Also, DICOM files for T1-weighted anatomical image should be included in another sub-directory and DICOM files for T2-weighted same slice image should be included in the other sub-directory.

Result

After the process is completed, you will find the converted files are in appropriate directories.



- Preprocessing of fMRI data and Univariate analysis for ROI definition

We rely on SPM5 for the preprocessings and univariate analysis.

Script

```
my_spm_run(MY_VAR)
```

Example of Batch files

```
myv_spm_run.m
my_script_spm_run.m
```

Preparation

The directories including NIFTI files for a participants should be in one directory. The anatomical T1 and T2 same slice images should be one directory (see bottom of the previous page).

The output of `convert_DICOM_to_NIFTI` is automatically organized correctly.

If you want to perform the first-level SPM analysis, you have to prepare SPM_mult_des_1.mat, SPM_mult_des_2.mat ... Each of these files corresponds to each runs. They should include three cell arrays of `durations`, `names`, `onsets`.

For the detail of the file, please see help of SPM about "multiple regressors."

Result

The outputs of preprocessings are saved in the same directories as the original NIFTI images. The outputs of first level analysis (e.g. SPM.mat etc.) are saved in an additional directory. Batch files for SPM is saved in the experimental directories for ascertainment of what has been done.

Note

If you will run each steps (e.g. realignment, coregistration, normalization) with separate scripts, please make sure that MY_VAR.prefix is correctly identified for each step of the preprocessing.

•Voxels (Spatial ROI) and Volumes (Temporal ROI) selection.

- Spatial ROI Definition

The pattern classification analysis often done with the prior voxel selection procedure. The script is to define the spatial ROIs.

Script

```
ROI_spatial(MY_VAR)
```

Example of Batch files

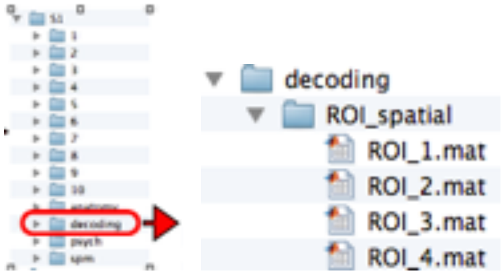
```
myv_ROI_spatial.m
my_script_ROI_spatial.m
```

Preparation

At least two SPM result files are required; the deformation function (y_deformation.mat) and t-map files (spmT_XXXX.hdr, spmT_XXXX.img).

Results

ROI definition files are saved sub-directory in the directory for decoding analysis.



Definition of ROI

You can choose the criterion for the prior voxel selection from 5 types.

1) Whole brain

All voxels in the brain is selected. The voxels out of the brain is discarded.

2) Anatomical Definition

The voxels that is in a certain brain areas defined by Anatomy Toolbox is selected. You can define the combination of multiple areas.

3) Definition with statistical values obtained from univariate analysis (SPM)

The voxels that has high t-values with SPM analysis are selected. The threshold of the height of the t-value should be defined. The definition can be whether t-value itself (e.g. voxels with t-score > 3.0) or number of voxels to be selected (e.g. top 200 voxels).

4) Intersection Combination of Anatomical definition and definition statistical values

The voxels that has highest t-values and in the defined anatomical ROI is selected. The anatomical ROI and the threshold for t-value should be defined.

- Temporal ROI Definition

The pattern classification analysis is performed with the fMRI volumes that is recorded in a certain timing (ex. 6-sec after participants performing task).

Script

```
ROI_temporal(MY_VAR)
```

Example of Batch files

```
myv_ROI_temporal.m
my_script_ROI_temporal.m
```

Preparation

The parameter files for decoding analysis should be prepared (Figure 9 and Example).

Result

ROI definition files are saved sub-directory in the directory for decoding analysis.

Leave two session out cross validation

If the leave-two session out cross-validation is needed, the additional definition files should be generated by using `ROI_temporal_leave_two_out(MY_VAR)`. In this case, the ROI definition files are saved in the sub-directories of the directory where the output of `ROI_temporal` is saved.

Definition of ROI

This script allows to define the volumes to be used for decoding. One volume or mean of multiple volumes are available.

Also, you can re-define the label by converging the two or more label into one. For example, when the participants moved their index or middle fingers of the right or left hand, the original model is two by two, and the decoding is multi-class classification problem. However, if you want to decode whether participants used whether right or left hand, you can converge the labels for right-index and right-middle fingers into one label.

•Decoding Analysis

- Pattern Classification

This script is designed to match the input and output across algorithms to make further analysis easier.
For detail, please see the other description file, “README_clsfy.”

Script

```
[errTable_tr,errTable_te,model] = clsfy(xtrain,ttrain,xtest,ttest,method)
```

Input

Classifier is trained with training data set (`xtrain` and `ttrain`) with specified method (`method`), and evaluated the performance with test data set (`xtest` and `ttest`).
`xtrain` and `xtest` should be a matrix whose size is [Number of Samples] × [Number of Features] and `ttrain` and `ttest` should be a matrix whose size is [Number of Samples] × 1.
Thus, `size(xtrain, 1)` should be equal to `size(ttrain, 1)`, `size(xtest, 1)` should be equal to `size(ttrain,1)` and `size(xtrain,2)` should be equal to `size(xtest,2)`

Output

The outputs `errTable_tr` and `errTable_te` are error tables for training and test data set. `model` contains other parameters of classifiers, which is used for further analysis.

- Leave-One-Run-Out Cross Validation

The cross validation for the performance evaluation is done. Before the classification, temporal normalization and HPF can be applied to the data (optional).

Script

```
cross_valid(MY_VAR)
cross_valid_leave_two_out(MY_VAR)
```

Example of Batch files

```
myv_cross_valid.m
my_script_cross_valid.m
myv_cross_valid_leave_two_out.m
my_script_cross_valid_leave_two_out.m
```

Preparation

The spatial ROI files and the temporal ROI files should be prepared. All combination of spatial and temporal ROIs are tested.

Result

Result files `decoding_res_1.mat`, `decoding_res_2.mat`... are saved in a sub-directory of decoding directory (Figure 9). Note that, the indexes of the results files are irrelevant to the indexes of the problem.

Leave Two Run Out Cross Validation

If you need to perform leave two run out cross validation, in order for parameter fitting, you also have to run `cross_valid_leave_two_out(MY_VAR)`. The results files of this is saved in the sub-directories named “leave_two_out”.

Parallel Processing

If you can use cluster machines and multiple MATLAB licenses, we recommend to use the option for parallel computing. Our parallel computing method is an pseudo-master-slave method. The divided problems are saved in files in one directory (master process). After the all problems are divided, you can run slaves as many as you like, by going to the current directory of the “master process,” and run “`parallel_process_slave.m`.” Slaves will keep solving the divided problems until all the problems are solved. Each problem contains, the decoding with one algorithm for one temporal ROI, for one spatial ROI, for one participant.

Tips for parallel processing:
・For our impression, multiple MATLAB sessions with “single core processing,” is faster than one MATLAB session with “multiple core processing.”

A known bug in parallel processing:
・ When two slaves tries to save “number.mat”, the file will corrupt and all the slaves will stuck. This rarely occurs. If you face the problem, please re-try from the unsolved problem with the earliest index. This can be done by discarding the corrupted `number.mat` in the parallel directory, and re-making it by

```
cd [parallel directory]; my_number = [X]-1; save number my_number
```

Replace `[parallel directory]` with the full path of the directory which contains `my_number.mat` and `parallel1.mat, parallel2.mat...` and replace `[X]` with the earliest index of the unsolved problem.

•Making and Saving Figures

- <u>Save Figures</u>

The information about classification performance is extracted from results files that is output of cross validation process. Then, many figures are saved as single .ps file.

Script files

Run `save_figures.m`, which processes the sub functions below.

Sub-functions

`save_figures_results_extract.m`
Extracting necessary information and combining it to one file
`save_figures_general_info.m`
Write cover page of figure file
`save_figures_compare_decoders.m`
Write figures for comparison between decoding algorithms.
`save_figures_compare_teROIs.m`
Write figures for comparison between volumes to be used for decoding.

Sub-sub-functions.

For Figure formatting we use
`add_return.m`
`area_list.m`

Example of Batch files

`myv_save_figures.m`
`my_script_save_figures.m`