run_cpm_sort_gender.m

Written by Suyeon Ju (suyeon.ju@yale.edu) on 1.25.22

Updated by Suyeon Ju on 4.1.22

Script location

- MRRC server:

  `/data23` → `mri_researchers` → `fredericks_data` → `shared` → `hcp_aging_analyses` → `hcp-a_cpm` → `run_cpm_sort_gender.m`

- Box:

  `Box` → `CPM_HCP-A` → `run_cpm_sort_gender.m`

How to run this script

Inputs:

- `subj_list_all` = cell array of .txt file names of all lists of subject IDs (can input more than one .txt file at once)

  - should be file name of .txt file that has list of all subject IDs [`'HCA#######'`]

  - i.e., `{'all_subjs_ravlt_hcp-a_2.txt', 'all_subjs_pcps_hcp-a_2.txt', 'all_subjs_nffi_hcp-a_2.txt'}`

- `behav_param_list` = cell array of behavioral parameters to be tested (can input more than one parameter at once)

  - i.e., `{'ravlt', 'pcps', 'nffi'}`

- `scan_type_list` = cell array containing one or more of the following (can input more than one scan type at once):

- - `'rfMRI_REST1_AP'`, `'rfMRI_REST1_PA'`, `'rfMRI_REST2_AP'`, `'rfMRI_REST2_PA'`, `'tfMRI_CARIT'`, `'tfMRI_FACENAME'`, `'tfMRI_VISMOTOR'`
  - i.e., `{'tfMRI_CARIT', 'tfMRI_FACENAME', 'tfMRI_VISMOTOR'}`

Outputs:

- `all_results_m` = cell array holding all male subjects' CPM results (`behav_param`, `scan_type`, `behav_scores_m`, `y_hat_m`, `corr_m`)

- `all_results_f` = cell array holding all female subjects' CPM results (`behav_param`, `scan_type`, `behav_scores_f`, `y_hat_f`, `corr_f`)

- For now, the code just prints the contents of the `all_results` cell arrays and resulting correlations between actual and predicted y (R and p value, for both male and female predictive models), but I will revise the code later to collect and save all CPM outputs (`y_predict` and `performance`)

Example command line:

```
run_cpm_sort_gender({'all_subjs_ravlt_hcp-a_2.txt', 'all_subjs_pcps_hcp-a_2.txt',
'all_subjs_nffi_hcp-a_2.txt'}, {'ravlt', 'pcps', 'nffi'}, {'tfMRI_CARIT',
'tfMRI_FACENAME', 'tfMRI_VISMOTOR'})
```

Pseudocode (what the script does, in a nutshell)

> All of the following steps are within a nested for-loop to allow for batch parameter runs (*outer loop* = *loops through list of selected behavioral parameters;* *inner loop* = *loops through list of scan types*)

step 1: create string array with all subj IDs from `subj_list_all`

step 2: create for loop where:

- iterative variable = all subj IDs

- `conn_mat_single` = connectivity matrix for each subj ID for specified scan_type; each conn_mat_single matrix is added in 3rd dimension to `conn_mat` (holds all conn mats across all subjs)

- `conn_mat` = compilation of all conn mats across all subjs

- `conn_subj_array` = collect subj IDs of all subjs in `conn_mat`

step 3: create separate cell arrays for male (`behav_scores_m`) and female (`behav_scores_f`) subjects that collect:

- all subj IDs in `conn_subj_array` (in col 1)

- corresponding behavioral parameter data pulled from `behav_param` (in col 2)

- corresponding gender of each subj (in col 3)

- corresponding connectivity matrix for each subj (in col 4)

step 4: extract male (`conn_mat_m`) and female (`conn_mat_f`) connectivity matrices from the `behav_scores` cell arrays (for ease of inputting the matrices into CPM in step 5)

step 5: call `cpm_main` function from constable's CPM matlab code separately for each gender (generates predictive model and outputs results for each gender separately!)

Example output:

```
>> run_cpm_sort_gender({'all_subjs_ravlt_hcp-a_2.txt'}, {'ravlt'},{'tfMRI_CARIT'})
ravlt
all_subjs_ravlt_hcp-a_2.txt
tfMRI_CARIT
Warning: The DATETIME data was created using format 'MM/dd/uuuu' but also matched 'dd/MM/uuuu'.
To avoid ambiguity, supply a datetime format using SETVAROPTS, e.g.
  opts = setvaropts(opts,varname,'InputFormat','MM/dd/uuuu');
> In matlab.io.internal.text.TableParser/readData (line 82)
  In matlab.io.internal.functions.ReadTableWithImportOptionsText/executeImpl (line 76)
  In matlab.io.internal.functions.ReadTableWithImportOptions/executeImpl (line 18)
  In matlab.io.internal.functions.ReadTableWithImportOptionsText/execute (line 122)
  In matlab.io.internal.functions.ReadTableWithImportOptions/execute (line 25)
  In matlab.io.internal.functions.ExecutableFunction/validateAndExecute (line 98)
  In matlab.io.ImportOptions/readtable (line 490)
  In run_cpm_sort_gender (line 75)
Warning: Data: 1 subjects have missing nodes. Please check your data.
> In cpm_check_errors (line 45)
  In cpm_main (line 50)
  In run_cpm_sort_gender (line 159)

# Running over 2 Folds.
Performing fold no. 1 2
Done.
M
    'ravlt'     'tfMRI_CARIT'     {243×4 cell}     [243×1 double]     [1×2 double]


ans =

    1     5

    0.1655    0.0097


# Running over 2 Folds.
Performing fold no. 1 2
Done.
F
    'ravlt'     'tfMRI_CARIT'     {323×4 cell}     [323×1 double]     [1×2 double]


ans =

    1     5

    0.2003    0.0003

>>
```