ISBI 2015 Longitudinal MS Lesion Segmentation Challenge

Evaluation Software Description 03/31/2015

Installation

The evaluation software was written in Matlab (source code included) and compiled with the Matlab compiler for Linux operating systems. Other operating system versions are available upon request. Use of the compiled version requires the 64bit 2013a (v81) runtime library, available from [1]. Extract the zip file, install, and set the MCRROOT variable in the `metrics4d.sh` script to the installed location. Please contact the conference organizers for any problems or bug reports.

Usage

For the compiled version, usage is as follows, where ref_mask is the ground truth, and seg_mask is the computed segmentation mask:

```
./metrics4d.sh seg_mask_time1.nii,seg_mask_time2.nii,…   ref_mask_time1.nii,ref_mask_time2.nii,…
```

Entering the program name without any arguments should provide brief usage instructions.

For the matlab version, usage is as follows:

```
Result=metrics4d('seg_mask_time1.nii,seg_mask_time2.nii,…','ref_mask_time1.nii,ref_mask_time2.nii,…
')
```

All files should be binary NIFTI (.nii extension) files, with two labels, 1 indicating lesions, and 0 indicating background. Note that although the nifti files are used as inputs to the evaluation software, header orientation information is ignored. Thus, the raw image spaces should be the same in the reference mask and the segmentation mask. Some testing is recommended to make sure that the output space is consistent with the masks within the training data.

Outputs

The following outputs are computed by the program in a comma separated format (many are described in [2]):

Dice Overlap – the ratio of twice the number overlapping voxels to the total number of voxels in each mask
Jaccard Overlap – the ratio of the number of overlapping voxels to the number of voxels in the union of each mask
PPV (positive predictive value) – the ratio of voxel-wise true positives to the sum of true and false positives
TPR (sensitivity, voxel based) – the ratio of voxel-wise true positives to the sum of true positives and false negatives
LTPR (lesion TPR based on lesion count) – the ratio of lesion-wise true positives to the sum of true positives and false negatives
LFPR (lesion FPR based on lesion count) – the ratio of lesion-wise false positives to the sum of false positives and true negatives
Volume Difference – absolute difference in volumes divided by the true volume
Surface Difference – average symmetric surface distance [2]
Segmentation Volume – total volume of segmentation mask for reference purposes
Manual Volume – total volume of reference mask for reference purposes
Volume Change Correlation – average linear correlation of changes in lesion volumes between successive time-points
New lesion detection TPR – ratio of number of new lesions detected to number of true new lesions
New lesion detection FPR – ratio of new lesions falsely detected to number of true new lesions

A lesion is considered to be detected if at least one voxel overlaps with a lesion voxel on the reference segmentation. If there are no new lesions between timepoints in the reference masks, the new lesion detection TPR will be NaN.

Evaluation

Algorithms will be ranked based on a combination of the metrics described above, along with overall lesion volume intraclass correlation. Even though a large number of metrics are computed, some may contribute little or not at all to the final ranking. For example, voxel-wise overlap metrics are generally regarded as highly variable measures of lesion segmentation performance (so will likely be given low weight for the rankings) but they still can be useful measures for characterizing the algorithm. The final formula for combining the metrics is being determined independently from the results from the competing algorithms, and will be provided prior to the release of the second test data set.

Although the challenge results are a useful resource to serve as an indicator of algorithm performance, given the limited data size and specific characteristics of this data set, they should not be considered as a sole and definitive benchmark.

References

[1] http://www.mathworks.com/products/compiler/mcr/

[2] E. Geremia, O. Clatz, B.H. Menze, E. Konukoglu, A. Criminisi, N. Ayache, “Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images,” 57:378-390, NeuroImage.