Model Validation and Automated Search in FMRI Path Analysis: A Fast Open-Source Tool for Structural Equation Modeling

Gang Chen[1], Daniel R Glen[1], Jason L Stein[2], Andreas S Meyer-Lindenberg[2], Ziad S Saad[1], Robert W Cox[1]
[1] *Scientific and Statistical Computing Core,* [2] *Unit for Systems Neuroscience in Psychiatry*
National Institute of Mental Health, NIH, DHHS, 9000 Rockville Pike, Bethesda, MD 20892
http://afni.nimh.nih.gov/sscc/gangc/PathAna.html

Overview: *1dSEM* in AFNI

- A path analysis program written in C (meaning that it is fast)
- Open source available for Unix+X11 systems (Linux, Mac OS X, Solaris, …)
- Independent of FMRI data analysis platform
- 2 modes
  - Model validation: confirmatory
  - Model search: exploratory
- 2 options in model search
  - Tree growth: hierarchical model search
  - Forest growth: search of all possible models
- Required input for model validation
  - Covariance / correlation matrix
  - Error variance at each region (optional)
  - Degrees of freedom
  - Region connections
- Required input for model search
  - Covariance / correlation matrix
  - Error variance at each region (optional)
  - Degrees of freedom
  - Desired, excluded, and searchable paths
- Performance with NEWUOA optimization [3]
  - Low computation cost
  - Among 4 published models fitted with LISREL: *1dSEM* validated 2; found better fits for other 2
  - Model search options outperform approaches with modification index or Lagrangian multiplier commonly adopted in SEM programs
- Caution on model search
  - Path analysis is essentially confirmatory
  - Incorporate prior knowledge into the search
  - Don't let data lead the search

Introduction

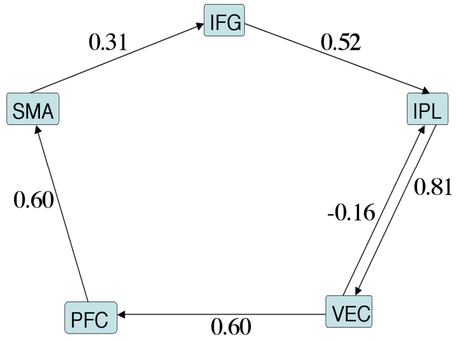
Connectivity modeling explores the hierarchical integration of brain regions under various psychological effects in FMRI data analysis. Two categories of connectivity analysis are usually employed in the process: **correlation analysis** is a data-driven, voxel-wise approach in which a seed region is tested for temporal correlation with the rest of the brain; **causal relation analysis** focuses on a few regions of interest and a testable model of the inter-regional correlations that is formulated with prior knowledge. Path analysis or **s**tructural **e**quation **m**odeling (SEM) is a theory-oriented or model-driven approach to causal modeling.

1dSEM, an SEM program implemented in AFNI and based on previous work on FMRI path analysis ([1], [2]), takes interregional covariances or correlations as input and estimates the connection strengths among the regions in the network. Instead of using adaptive simulated annealing [2], we adopted a nonlinear optimization scheme, NEWUOA [3], for minimizing the maximum likelihood cost function. This method allows us achieve high efficiency and speed. We also modified the original NEWUOA to allow constraints and to search along multiple paths so that it is less likely to get trapped at a local optimum. Compared to LISREL's optimization method, our method gave equal or better fits. In four published models, our method confirmed 2 models but generated a better fit in the other 2 cases.

The commonly used algorithm in model search in SEM is via modification index or Lagrangian multiplier in packages such as LISREL. However, we employed two alternatives to model modification: tree growth and forest growth. In tree growth, an extra path grows as a new branch on the previous model tree, and the new branch is chosen with the best fit among all possible paths. On the other hand, forest growth is a brute-force method that simply searches through the forest of all possible networks with each specific number of paths and selects the network with the best fit.

Model Validation

Data from [1] were used to test *1dSEM* with a model of 5 regions in the articulatory system.



Command line

```
1dSEM -theta thetas0.1D  -C corr.1D  -psi psi.1D  -DF 30
          -limits -2 2
```

Where

-theta specifies the model to be tested, from text file *thetas0.1D*

```
#     VEC    PFC    SMA    IFG   IPL
VEC    0      0      0      0     1
PFC    1      0      0      0     0
SMA    0      1      0      0     0
IFG    0      0      1      0     0
IPL    1      0      0      1     0
```

-C specifies correlation/covariance matrix from text file *corr.1D*

```
1       0       0       0       0
0.661   1       0       0       0
0.525   0.66    1       0       0
0.486   0.507   0.437   1       0
0.731   0.63    0.558   0.517   1
```

-psi specifies residual error variances from text file *psi.1D*

```
0.825
0.868
0.87
0.881
0.851
```

-limits specifies the path coefficient range,

and *-DF* specifies the degrees of freedom.

Output on screen (runtime: ~ 2 seconds)

```
….
++ Cost is 0.429081, chi-square (9) = 12.4434 (P=0.19)
AIC = 24.44, RMSEA = 0.31, CFI = 0.94, GFI = 0.83
Connection coefficients matrix: 5 x 5
#        VEC      PFC      SMA      IFG      IPL
VEC    0.0000   0.0000   0.0000   0.0000   0.8076
PFC    0.5974   0.0000   0.0000   0.0000   0.0000
SMA    0.0000   0.5961   0.0000   0.0000   0.0000
IFG    0.0000   0.0000   0.3144   0.0000   0.0000
IPL   -0.1589   0.0000   0.0000   0.5231   0.0000
```

The above result matches the model in [1] estimated in LISREL.

For 3 frontoparietal working memory networks in other published data, *1dSEM* confirms only one model and gives differing path coefficients (even differing sign in some paths) for the other 2 models but with a better fit than the LISREL method. (Details available on request.)

References

[1] **ET Bullmore**, *et al.* How good is good enough in path analysis of FMRI data? NeuroImage **11**:289-301 (2000).

[2] **JL Stein**, *et al.*, A validated network of effective amygdala connectivity. NeuroImage *in press* (e-pub 28 Mar 2007).

[3] **MJD Powell**, "The NEWUOA software for unconstrained optimization without derivatives", Technical report DAMTP 2004/NA08, Cambridge University Numerical Analysis Group (http://www.damtp.cam.ac.uk/user/na/NA_papers/NA2004_08.pdf).

See Also

Poster Tue AM **#113** by Pezawas *et al.*
Poster Tue PM **#336** by Saad *et al.*
Poster Wed PM **#300** by Cox *et al.*

Model Search

Again, the data from [1] were used to search an optimal model for each fixed number of paths in terms of various fit indices using *1dSEM* with a model of 5 regions in the articulatory system

Tree Growth

Search for optimal model by growing a model for one additional coefficient from the previous model for $n-1$ coefficients. If the initial theta matrix has no required coefficients, the initial model will grow from the best model for a single coefficient

Command line

```
1dSEM -tree_growth -theta thetas1.1D -C corr.1D -psi psi.1D  \
          -DF 30 -max_paths 6 -limits -2 2
```

Where

-theta specifies paths to be included (`1`), excluded (`0`), and searchable (`2`), from text file *thetas1.1D*

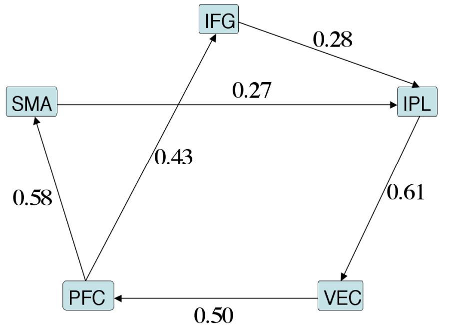
```
#     VEC     PFC    SMA    IFG    IPL
VEC    0       0      0      0      1
PFC    1       0      0      0      2
SMA    0       2      0      0      0
IFG    2       2      2      0      0
IPL    2       0      2      2      0
```

-max_paths specifies the maximum number of paths in the search

-C, *-psi*, *-DF* , and *-limits* remain the same as in Model Validation.

Output on screen (runtime: ~ 2 seconds)

```
….
++ cost = 0.38202, chisq(9) = 11.0786 (P=0.27), ntheta = 6
++ parsimonious fit index = 0.537 AIC = 23.0786
RMSEA = 0.24, CFI = 0.96, GFI = 0.85
Connection coefficients matrix: 5 x 5
#        VEC      PFC      SMA      IFG      IPL
VEC    0.0000   0.0000   0.0000   0.0000   0.6059
PFC    0.5022   0.0000   0.0000   0.0000   0.0000
SMA    0.0000   0.5813   0.0000   0.0000   0.0000
IFG    0.0000   0.4258   0.0000   0.0000   0.0000
IPL    0.0000   0.0000   0.2723   0.2776   0.0000
…
```



The above result corresponds to the one in [1] estimated using LISREL with tree growth via modification index.

Forest Growth

Search over all possible path combinations by comparing models at incrementally increasing number of path coefficients.

Command line

```
1dSEM  -forest_growth  -theta thetas1.1D  -C corr.1D \
          -psi psi.1D  -DF 30  -max_paths 6  -limits -2 2
```

Forest growth gives the same optimal network of 6 paths (runtime: ~ 5 seconds) with the data in the articulatory system [1] as tree growth, although models with other path numbers differ.

Summary

Structural equation modeling requires robust optimization algorithms, and model searches can vary dramatically depending upon search methods and model specifications as well. The optimization scheme presented here gives better fits, and the forest growth method allows all possible models to be explored. The ability of *1dSEM* to allow the user to require or exclude various paths gives the flexibility to interject what is known about the modeled system.