PREDICTION OF RECOVERY FROM MILD TBI USING GENETIC PROGRAMMING ANALYSIS OF DTI DATA

Richard Watts[1], Margaret J. Eppstein[2], Alex Thomas[3], Joshua P. Nickerson[1], Hugh Garavan[4], Trevor Andrews[1,5], Christopher G. Filippi[6], and Kalev Freeman[3]

[1]Department of Radiology, University of Vermont College of Medicine, Burlington, VT, United States, [2]Department of Computer Science, University of Vermont, Burlington, VT, United States, [3]Department of Surgery, University of Vermont College of Medicine, Burlington, VT, United States, [4]Department of Psychiatry, University of Vermont College of Medicine, Burlington, VT, United States, [5]Philips Healthcare, Cleveland, OH, United States, [6]University of Vermont College of Medicine, Department of Neurology, Burlington, VT, United States

Purpose

To predict recovery from mild traumatic brain injury (mTBI) using a novel application of genetic programming (GP) to longitudinal DTI data acquired in the first week post-injury. GP is a powerful technique that implicitly combines feature selection, model identification and parameter estimation, and is able to discover non-linear relationships between parameters.

Methods

In this IRB-approved longitudinal study, 20 mTBI patients were scanned within 72 hours of injury (46.5±22 hours) and at follow-up one week later. 16 control subjects were similarly imaged at two time-points separated by one week. Data was acquired using a Philips 3T Achieva TX scanner and an 8-channel head coil. Diffusion weighted images were acquired with 46 uniformly distributed directions at b=1000s/mm^2 and 6 with no diffusion weighting, at 2mm isotropic resolution. Eddy current correction, tensor fitting and spatial normalization were performed using FSL. Fractional anisotropy values were extracted from eleven *a priori* regions of interest (ROIs) most commonly reported in the TBI literature(1); the splenium (SCC), body (BCC) and genu (GCC) of the corpus callosum, and the left and right posterior limbs of the internal capsule (PLIC), uncinate fasciculus (UF), corona radiata (CR) and corticospinal tract (CST).

Symbolic regression was performed using the GP package Eureqa(2) to predict the sum of post-concussive symptoms at follow-up (SS_2) based on baseline symptoms (SS_1) and changes in FA values in each region. In GP, the evolving genotypes are expression trees representing functions involving the symbols {+,-,×,/,>} (Figure 1). Mutation is accomplished by changing the function or replacing random subtrees. Recombination is accomplished by exchanging subtrees between candidate solutions. Data from a random set of 18 subjects (10 mTBI cases and 8 controls) was used for training; the remaining 18 subjects used for validation. The fitness of a particular expression is based on R^2 of predicted vs. observed SS_2 on the training set.

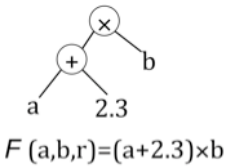

Figure 1. An example of an expression tree and the function it represents.

Results

As expected, some of the variability in SS_2 was attributable to variability in SS_1 (R^2=0.439). However, we found several expressions using the GP that were much more highly correlated with SS_2. For example:

$$SS_{2,predicted} = SS_1\left(49\Delta FA_{SCC} + 5443\Delta FA_{SCC}^2\right) \qquad \textbf{Equation 1}$$

where ΔFA_{SCC} represents the longitudinal change in FA in the splenium of corpus callosum (SCC). The predicted SS_2 using equation (1) was highly positively correlated to the observed SS_2 (Figure 2). For the validation data (n=18) R^2=0.948 while *post-hoc* analysis (SPSS 22) of the entire group (n=36) gave R^2=0.916 (n=36).

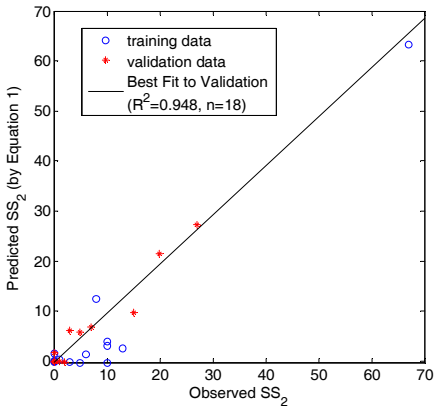

Figure 2. The predictions of Equation 1 are highly correlated with the observed symptom scores at follow-up.

Discussion and Conclusions

Imaging data is conventionally analyzed using linear models. The use of GP allows the model itself to be determined by the data, allowing novel solutions to be automatically discovered. Eureqa provides the user with a range of solutions that are non-dominated with respect to fitness and complexity of the expression tree (Figure 3). A more complex expression will generally be able to fit the training data more accurately, but is more susceptible to overfitting.

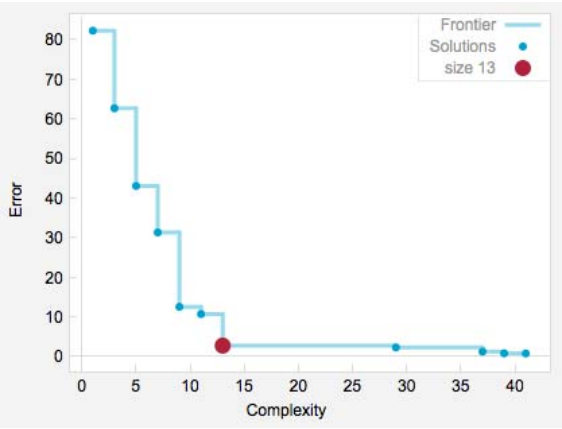

Figure 3. Decreased error with increased complexity. Highly complex solutions fit the observed data well, but generalize poorly to new data.

The form of the Equation 1 is encouraging; the SCC was identified as the most predictive region (feature selection), consistent with prior literature and conventional linear regression; controls and patients who are asymptomatic at the first time point (SS_1=0) are unlikely to subsequently develop symptoms; unchanged FA in SCC is associated with full recovery while large changes predict poor outcome. The quadratic function of ΔFA_{SCC} can be regarded as a polynomial expansion.

While the sample size in the current study is too small to allow proper cross-validation, these results are suggest that GP may be useful to elicit non-linear relationships between imaging and clinical findings.

References

1. Hulkower MB, Poliak DB, Rosenbaum SB, Zimmerman ME, Lipton ML. A Decade of DTI in Traumatic Brain Injury: 10 Years and 100 Articles Later. AJNR American journal of neuroradiology 2013.

2. Schmidt M, Lipson H. Distilling Free-Form Natural Laws from Experimental Data. Science 2009;324(5923):81-85.