Regression into early adulthood: a data-driven perspective of NIH longitudinal pediatric DTI study

Yasheng Chen[1], Hongtu Zhu[1], Hongyu An[1], Dinggang Shen[1], and Weili Lin[1]
[1]*UNC-CH, Chapel Hill, NC, United States*

Introduction

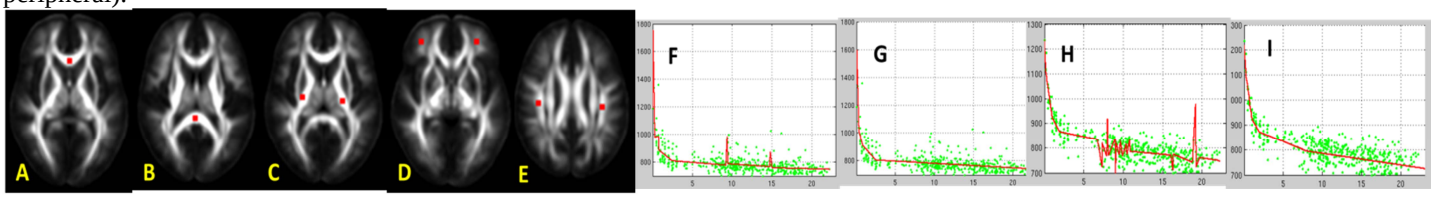
Regression analysis on diffusion tensor imaging (DTI) aims for the statistical inference on a certain physiological or pathophysiological process through modeling the apparent changes of a series of diffusion related parameters such as fractional anisotropy or mean diffusivity (MD). In particular, the onset of a physiological event may perturb the growth trajectory from its original velocity. As in brain maturation, nonlinear growth trajectories were observed almost in all age spans [1-3]. Most of the current brain developmental studies were based upon a global parametric model such as nonlinear polynomials. These approaches may not be able to capture subtle local temporal changes of DTI parameters and the physiological meanings of high order polynomial terms may be elusive. We propose to analyze complex brain growth patterns in large scale datasets from NIH longitudinal DTI pediatric study with a data-driven regression method combined with longitudinal statistical analysis for a better inference on the growth trajectory of brain development from birth to young adulthood.

Methods

Our data-driven analysis consists of two major steps: 1) deriving an initial growth trajectory with a semi-parametric regression analysis and 2) pruning the coefficients of the obtained growth trajectory through a series of longitudinal statistical tests. We approximate the complex growth trajectory under a free-knot B-spline fitting scheme [4]. The growth data are given as pairs of $\{x_i, y_i\}$ with observation times $0 \le x_1 \le x_2, \ldots \le x_n$. The measurements is modeled as $y_i = f(x_i) + e_i$ (e_i is the noise associated with the i^{th} measurement). The function f is inferred using a spline model with a knot sequence, $t_1=t_2=\ldots t_k<t_{k+1}<t_{k+2}<..<t_n=t_{n+1}=..t_{n+k}$ (order k). Free-knot spline enables knot relocation and knot removal during the fitting. The B-spline model was initialized with a very large number of knots to capture all important temporal features presented within the data. The side effect is that this will always lead to an over-fitted model. The knots were ranked according to the jumps of the first discontinuous derivative of the two neighboring spans. If removal of the least important knot leads to an unacceptable fitting, the previous knot sequence will be determined as the final output. After knot removal, the remaining knots will be repositioned to minimize the fitting error. The second step performs statistical testing on the significance of the identified knots on the growth trajectory with quasi-least squares (QLS) method. QLS is an alternative estimate of the correlation parameters within the generalized estimating equations (GEE) frame work, which guarantees a consistent estimate of the correlation parameter and a positive definite correlation matrix [5]. We used linear B-spline models to generate an over-fitted piecewise linear growth trajectory in the first step as: $f(t) = a_0+a_1*t+a_2*max(t-b_1, 0) + a_3*max(t-b_2,0)+\ldots$ In second step, the most insignificant knots were removed one at a time until all remaining knots making significant contributions to the growth trajectory. With the final piecewise linear representation, the identified significant knots may reflect the occurrence of physiological events causing the transition in growth velocities of DTI.

Results

A total of 458 longitudinal DTI datasets from 274 subjects covering the life span from birth to 22 years of age were included in the study. All DTI images were registered towards the ICBM DTI template (A to E). We selected a few white matter regions respectively in central (A for genu, B for splenium and C for posterior limbs of internal capsule) and peripheral white matter regions (D for frontal white matter and E for superior longitudinal fasciculus). The original growth trajectories for both central and peripheral white matter generated by knot-free spline fitting had spikes or oscillations from the over-fitted model (F for central and H for peripheral). After removal of insignificant knots with QLS, smoother growth trajectories were obtained (G for central and I for peripheral).



Discussion and Conclusions

In this work, we have proposed a powerful approach for regression analysis of large longitudinal datasets from NIH pediatric DTI brain developmental study. Through the combination of the greater flexibility of the free-knot B-spline fitting with the QLS longitudinal statistical analysis, we are able to delineate the complex process of brain growth from birth to early adulthood into a series of linear spans so that physiological inferences on brain growth can be attributed to the transitions of growth velocities.

References

[1] Lebel, C., et al., 2008. Microstructural maturation of the human brain from childhood to adulthood. Neuroimage 40, 1044-1055. [2] Chen, Y., et al., 2011. Longitudinal regression analysis of spatial-temporal growth patterns of geometrical diffusion measures in early postnatal brain development with diffusion tensor imaging. Neuroimage 58, 993-1005. [3] McKinstry, R.C., et al., 2002. Radial organization of developing preterm human cerebral cortex revealed by non-invasive water diffusion anisotropy MRI. Cereb Cortex 12, 1237-1243. [4] Schwetlick, H., et al., 1995. Least squares approximation by splines with free knots. [5] Shults,J., et al., Analysis of serially correlated data using quasi-least squares, Vol. 54, No. 4 (Dec., 1998), pp. 1622-1630.