Clustering and Partitioning of Functional Data for Understanding the Heterogeneity in Psychiatric Diagnosis

Eva Petkova[1], Thaddeus Tarpey[2]

[1] Department of Child and Adolescent Psychiatry, New York University, USA
[2] Department of Mathematics, Wright State University, USA

In the past two centuries progress has been made in many medical areas with the development of objective tests and understanding of the cause (etiology) and mechanism of diseases (pathophysiology) . Unfortunately, the attempts to understand causes of psychiatric illness by studying genetic or brain functions did not yield the hoped for results and psychiatry still depends on non-objective descriptions of syndromes. Some psychotropic drugs can specifically induce remission, suggesting an experimental approach to understanding the pathophysiological processes. This approach requires evidence that the medication directly affects the disease mechanism, rather than exerting benefit by non-specific compensation. For each particular patient who remits during drug treatment, it is unknown if s/he benefited from a specific pharmacological action or from a non-specific placebo effect or spontaneous remission. If psychiatric illness consists of distinct diagnostic categories (e.g., subjects who's illness has a mechanism that can be affected by a specific pharmacologic agent, vs. those illness is not affected by this pharmacologic agent), then we should be able to cluster responders to the treatment into distinct classes based on, for example, course of the severity of symptoms over time, medical history and other measures. This would allow to isolate a purer subset of the subjects with more homogeneous response to the drug and, possibly, with similar mechanism of the disease. Importantly, even if there are no distinct diagnostic categories, we will still want to isolate subjects with similar response to the drug, i.e. similar course of symptom decrease over time, in the hopes that homogeneity in the effect of drug on the symptoms reflects analogous pathophysiology.
In this paper we describe a strategy for studying the heterogeneity of functional (over time) response to treatment. *B*-splines are employed to model symptoms' course over time for dimension reduction of the functional data. Latent mixed effects model [1] is used to test for presence of distinct clusters (diagnostic categories); covariates are used both in the model for the functional measurements and in the latent class model. When there is no evidence for more than one cluster, a parametric k-means and a parametric principal points algorithms are offered for optimal partitioning of the homogeneous distribution of the functional response into K parts [2]. The strategy is applied to functional outcomes from a depression treatment study with the goal of identifying subjects with specific response to the experimental pharmacologic agent.

References

[1] Lu, Y (2007) *A Random Effects Mixture Model for Classifying Functional Data*. Doctoral Dissertation, Columbia University, New York, NY

[2] Tarpey, T (2007) Linear transformations and the k-means clustering algorithm: applications to clustering curves. *The American Statistician*, Vol 61, pp 34-40.