A Non-Parametric Bayesian Change-Point Detection Method for Novice Teenage Driving Risk

Qing Li
Assistant professor, Iowa State University, qlijane@iastate.edu

Feng Guo
Associate professor, Virginia Tech Transportation Institute, feng.guo@vt.edu

Inyoung Kim
Associate professor, Virginia Tech, inyoungk@vt.edu

The driving risk during the initial period after licensure for novice teenage drivers is typically the highest but decreases rapidly right after. The change-point of driving risk is a critical parameter for evaluating teenage driving risk, which also varies substantially among drivers. This research proposes a non-parametric Bayesian method to detect the change-points of the intensity rates in the recurrent-event context and cluster subjects by their change-points. We assume that the event counts follow a non-homogeneous Poisson process with piece-wise-constant intensity functions. We propose a Dirichlet process mixture model allowing change-points to vary among subjects, while the change-points are assigned a Dirichlet process prior. The intensity rates are subject specific. A Markov chain Monte Carlo algorithm is developed to sample from the posterior distributions. The proposed approach automatically clusters subjects based on the change-points without the need to specify the number of latent clusters or conduct the model selection procedure. In the simulation study, we examine the model performance and also compare it with a Bayesian finite mixture model with predefined number of latent classes. The simulation study suggests that the proposed method is flexible and not seriously affected by changes in the model assumptions. It outperforms the Bayesian finite mixture model in several aspects, including detecting the correct number of clusters, assigning subjects to the correct cluster, and computing efficiency. We also apply the method to the Naturalist Teenage Driving Study, which continuously recorded the driving data of 42 novice teenage drivers for 18 months using advanced in-vehicle instrumentation. Application to the Naturalistic Teenage Driving Study identifies three distinct clusters with change-points at 60.42, 74.94, and 89.23 hours hours of driving after first licensure respectively. The intensity rates and the pattern of change also differ substantially among clusters. The results of this research provide more insight in teenagers' driving behaviour and will be critical to improve young drivers' safety education and parent management programs, as well as provide crucial reference for the Graduated Driver Licensing regulations to encourage safer driving.

Keywords: Naturalistic Teenage Driving Study, Cluster, Dirichlet Process Mixture Model, Non-Homogeneous Poisson Process