Statistical Learning with Time Series Dependence: An Application to Scoring Sleep in Mice

Abraham Wyner

We develop methodology which combines statistical learning methods with generalized Markov models, thereby enhancing the former to account for time series dependence. Our methodology can accommodate general and long-term time dependencies in an easily estimable and computationally tractable fashion. We apply our methodology to scoring sleep in mice. As currently-used methods are expensive, invasive, and labor intensive, there is considerable interest in high-throughput automated systems. Previous efforts have been able to differentiate sleep from wakefulness, but they are unable to differentiate the rare and important state of REM sleep from non-REM sleep. Key difficulties in detecting REM are that (i) REM is much rarer than non-REM and wakefulness, (ii) REM looks similar to non-REM in terms of the observed covariates, (iii) the data are noisy, and (iv) the data contain strong time dependence structures crucial for differentiating REM from non-REM. Our new approach (i) shows improved differentiation of REM from non-REM sleep and (ii) accurately estimates aggregate quantities of sleep in our application to video-based sleep scoring of mice.