

Automated Discovery of Multivariate Associations in Large Time-Varying data sets: a Healthcare Network Application

Narine Manukyan, Margaret J. Eppstein, Jeffrey D. Horbar, Kathleen A. Leahy, Michael J. Kenny, Shreya Mukherjee and Donna M. Rizzo

We introduce a new method for exploratory analysis of large data sets with time-varying features, where the aim is to automatically discover novel relationships between features (over some time period) that are predictive of any of a number of time-varying outcomes (over some other time period). Using a genetic algorithm, we co-evolve (i) a subset of predictive features, (ii) which attribute will be predicted (iii) the time period over which to assess the predictive features, and (iv) the time period over which to assess the predicted attribute. After validating the method on 15 synthetic test problems, we used the approach for exploratory analysis of a large healthcare network data set. We discovered a strong association, with 100% sensitivity, between hospital participation in multi-institutional quality improvement collaboratives during or before 2002, and changes in the risk-adjusted rates of mortality and morbidity observed after a 1-2 year lag. The results provide indirect evidence that these quality improvement collaboratives may have had the desired effect of improving health care practices at participating hospitals. The proposed approach is a potentially powerful and general tool for exploratory analysis of a wide range of time-series data sets.