

Using Statistical Significance and Prediction to Test Long/Short Term Public Services and Patients' Cohorts: A Case Study in Scotland

Authors : Raptis Sotirios

Abstract : Health and social care (HSc) services planning and scheduling are facing unprecedented challenges due to the pandemic pressure and also suffer from unplanned spending that is negatively impacted by the global financial crisis. Data-driven can help to improve policies, plan and design services provision schedules using algorithms assist healthcare managers' to face unexpected demands using fewer resources. The paper discusses services packing using statistical significance tests and machine learning (ML) to evaluate demands similarity and coupling. This is achieved by predicting the range of the demand (class) using ML methods such as CART, random forests (RF), and logistic regression (LGR). The significance tests Chi-Squared test and Student test are used on data over a 39 years span for which HSc services data exist for services delivered in Scotland. The demands are probabilistically associated through statistical hypotheses that assume that the target service's demands are statistically dependent on other demands as a NULL hypothesis. This linkage can be confirmed or not by the data. Complementarily, ML methods are used to linearly predict the above target demands from the statistically found associations and extend the linear dependence of the target's demand to independent demands forming, thus groups of services. Statistical tests confirm ML couplings making the prediction also statistically meaningful and prove that a target service can be matched reliably to other services, and ML shows these indicated relationships can also be linear ones. Zero paddings were used for missing years records and illustrated better such relationships both for limited years and in the entire span offering long term data visualizations while limited years groups explained how well patients numbers can be related in short periods or can change over time as opposed to behaviors across more years. The prediction performance of the associations is measured using Receiver Operating Characteristic(ROC) AUC and ACC metrics as well as the statistical tests, Chi-Squared and Student. Co-plots and comparison tables for RF, CART, and LGR as well as p-values and Information Exchange(IE), are provided showing the specific behavior of the ML and of the statistical tests and the behavior using different learning ratios. The impact of k-NN and cross-correlation and C-Means first groupings is also studied over limited years and the entire span. It was found that CART was generally behind RF and LGR, but in some interesting cases, LGR reached an AUC=0 falling below CART, while the ACC was as high as 0.912, showing that ML methods can be confused padding or by data irregularities or outliers. On average, 3 linear predictors were sufficient, LGR was found competing RF well, and CART followed with the same performance at higher learning ratios. Services were packed only if when significance level(p-value) of their association coefficient was more than 0.05. Social factors relationships were observed between home care services and treatment of old people, birth weights, alcoholism, drug abuse, and emergency admissions. The work found that different HSc services can be well packed as plans of limited years, across various services sectors, learning configurations, as confirmed using statistical hypotheses.

Keywords : class, cohorts, data frames, grouping, prediction, prob-ability, services

Conference Title : ICASSS 2021 : International Conference on Applied Statistics and Statistical Science

Conference Location : Athens, Greece

Conference Dates : April 08-09, 2021