Title: General Insurance Claims Modelling with Factor Collapsing and Bayesian Model Averaging

Authors: Sen Hu, Adrian O'Hagan, Brendan Murphy

Abstract: Insurance product pricing involves analysis of past insurance claims data as well as different properties of the insured objects and the corresponding policy holders, and generalised linear models (GLMs) have become the industry's standard approach for pricing and modelling risks of this nature. However, as typically implemented, the GLM approach utilises a single "best" model on which pricing outcomes and loss predictions are based, which can involve subjective decisions as to the inclusion or exclusion of variables that are borderline significant. An additional characteristic of most general insurance datasets is the presence of many categorical variables, each with multiple levels, which can adversely affect the parsimony of the model and the interpretability and communicability of its results. Particularly, not all levels of each factor variable may be required or statistically significant, and rather some subsets of the factor levels may be merged to give a smaller overall number of levels. This problem is more obvious when the number of levels within the factor is high. We propose a method for assessing the optimal manner to collapse a factor with many levels into one with a smaller number of levels, then using BMA to blend model parameters or predictions from all reasonably good models arising under every possible collapsed form of the factor in question. In this way, the parsimony of the models and the interpretability of their results will also be improved. This method will be particularly computationally intensive considering the number of factors being collapsed as well as the possibly large number of levels within factors. Hence a stochastic optimisation search is proposed to find the best few collapsing cases across the model space before BMA is used.