Enhanced Variable Selection Algorithm for Handling Overdispersion in Claims Frequency Modeling with Mixture of Poisson Regression Using Telematics Car Driving Data

Jennifer S. K. Chan[1], S. T. Boris Choy[2], Udi Makov[3], Ariel Shamir[4] and Vered Shapovalov[4]

1- School of Mathematics and Statistics, The University of Sydney
2- Discipline of Business Analytics, The University of Sydney
3- Actuarial Resreach Center, University of Haifa
4- Efi Arazi School of Computer Science, Reichman University, Herzliya

Abstract:

Automobile insurance pricing often relies on Poisson regression models to predict the number of claims, yet this assumption can be challenging to justify due to overdispersion. Alternative modeling approaches, such as zero-inflated models, are often considered to address this issue, as they represent special cases of finite mixture distributions. However, implementing finite mixture regression modeling with telematics data presents computational challenges, particularly in selecting variables to achieve predictive power without overfitting. This paper proposes an algorithm for variable selection in the presence of a large number of covariates by generating sub-samples corresponding to each component of the Poisson mixture. Variable selection is applied following the enhancement of the Poisson assumption by controlling the number of zero claims. The algorithm's performance is evaluated using out-of-sample AUC (Area Under the Curve), a measure of predictive power in machine learning. Application of the algorithm with claim history and telematics data reveals improved AUC compared to alternative Poisson regression algorithms. This approach facilitates more accurate pricing in the context of telematics data usage in automobile insurance.

Keywords: mixture Poisson regression; variable selection; telematics

References:

-Bermúdez, Lluis, Dimitris Karlis, and Isabel Morillo. 2020. Modeling Unobserved Heterogeneity in Claim Counts Using Finite Mixture Models. Risks 8: 10.
-Brown, Garfield O., Steve Brooks, and Winston Buckley. 2015. Experience rating with Poisson mixtures. Annals of Actuarial Science 9: 304–21.
- Guillen, Montserrat, Jens Perch Nielsen, Mercedes Ayuso, and Ana M. Pérez-Marín. 2019. The use of telematics devices to improve automobile insurance rates. Risk Analysis 39: 662–72
- Verbelen, Roel, Katrien Antonio, and Gerda Claeskens. 2018. Unravelling the predictive power of telematics data in car insurance pricing. Journal of the Royal Statistical Society: Series C (Applied Statistics) 67: 1275–304.