Chapter 94
A Data Mining Approach for Risk Assessment in Car Insurance:
Evidence from Montenegro

Ljiljana Kašćelan
University of Montenegro, Montenegro

Vladimir Kašćelan
University of Montenegro,, Montenegro

Milijana Novović-Burić
University of Montenegro, Montenegro

ABSTRACT

This paper has proposed a data mining approach for risk assessment in car insurance. Standard methods imply classification of policies to great number of tariff classes and assessment of risk on basis of them. With application of data mining techniques, it is possible to get functional dependencies between the level of risk and risk factors as well as better results in predictions. On the case study data it has been proved that data mining techniques can, with better accuracy than the standard methods, predict claim sizes and occurrence of claims, and this represents the basis for calculation of net risk premium and risk classification. This paper, also, discusses advantages of data mining methods compared to standard methods for risk assessment in car insurance, as well as the specificities of the obtained results due to small insurance market, such is the one in Montenegro.

1. INTRODUCTION

Insurance companies nowadays operate in conditions of growing competition. The aspiration for accelerated market growth implies attracting of greater number of users, and consequently the risk of losses is becoming higher. Decision making in such conditions requires as accurate evaluation of policy risks as possible, and defining of competitive premiums in order to reach the planned profit.

The majority of insurance companies keep the data on history of its operations in a data warehouse. These huge quantities of data are hiding very important information, which could contribute to easier decision making and risk assessment in car insurance. Data mining is capable of extracting this important information and it can also justify the investments of insurance companies in data.

This paper presents the possibilities, advantages and disadvantages of risk assessment in car insurance, with application of data mining techniques such as clustering, regression and classification trees. The second section provides the review of papers dealing with similar issues. Third section defines the concept of risk in car insurance and discusses standard methods for risk assessment and their shortcomings. Data mining techniques, used in this paper for risk assessment, are explained in this section, too. We have presented the capabilities of these techniques on data of the insurance company Sava Montenegro in section four. We start with a complete data set which is clustered to homogenous clusters i.e. clusters with similar amounts of claims. Expected claim sizes for identified clusters are evaluated with linear regression. For calculation of net risk premium, besides the amount, also the probability of claim occurrence is important, and in this paper it is evaluated with logistic regression. In this section we also suggested the decision tree technique for prediction of the policy risk level. In section five we discussed the obtained results, advantages and disadvantages of the applied data mining methods for assessment of risk on a small insurance market such is the one in Montenegro. Conclusions and future researches are discussed in the last section.

2. REVIEW OF LITERATURE

Aggregate claims for a homogeneous insurance portfolio have long been estimated using pure algorithmic methods (Chain-Ladder, Bornhuetter & Ferguson and Poisson) or simple stochastic methods (Generalized linear models, Bayesian, Distributional, Bootstrap method, and other) (Wuthrich & Merz, 2008; De Jong, & Heller 2008). Algorithmic, distribution-free methods use mechanical technics (run-off triangle) to predict claim reserves. This understanding does not allow for the quantification of the uncertainties in these predictions. Uncertainties can only be determined if we have an underlying stochastic model on which the prediction algorithms can be based. Some recent studies suggest improvements for the existing stochastic models (Björkwall, Hössjer, Ohlsson & Verrall, 2011; Brillinger, 2012; Zhang, Dukic & Guszcza, 2012).

For micro-level (level of individual claims), recent studies have perceived that a mixed discrete-continuous model may be appropriate to estimate claims and risk in insurance data (Christmann, 2004; Heller, Stasinopoulos & Rigby, 2006; Parnitzke, 2008; Bortoluzzo, Claro, Caetano & Artes, 2011; Huo,Wang, & Yang, 2013). According to Parnitzke (2008), the model explicitly specifies a logit-linear model for the occurrence of a claim (i.e. claim probability) and linear model for the mean claim size. Generalized linear models and more flexible Tweedie's compound Poisson model are often used to construct insurance tariffs (Smyth & Jorgensen, 2002). However, even this more general models still can

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/a-data-mining-approach-for-risk-assessment-in-car-insurance/150255

Related Content

Here Be Dragons: Mapping Student Responsibility in Learning Analytics
Paul Prinslooand Sharon Slade (2016). *Big Data: Concepts, Methodologies, Tools, and Applications (pp. 1717-1735).*
www.irma-international.org/chapter/here-be-dragons/150238

Sentimental Analysis Tools
Sunil M. E.and Vinay S. (2019). *Extracting Knowledge From Opinion Mining (pp. 204-231).*
www.irma-international.org/chapter/sentimental-analysis-tools/211560

Classification Of 3G Mobile Phone Customers
Ankur Jain, Lalit Wangikar, Martin Ahrens, Ranjan Rao, Suddha Sattwa Kunduand Sutirtha Ghosh (2007). *International Journal of Data Warehousing and Mining (pp. 22-31).*
www.irma-international.org/article/classification-mobile-phone-customers/1782

Multitarget Classifiers for Mining in Bioinformatics
Diego Liberati (2009). *Handbook of Research on Text and Web Mining Technologies (pp. 684-693).*
www.irma-international.org/chapter/multitarget-classifiers-mining-bioinformatics/21751

The Dynamics of Content Popularity in Social Media
Symeon Papadopoulos, Athena Vakaliand Ioannis Kompatsiaris (2010). *International Journal of Data Warehousing and Mining (pp. 20-37).*
www.irma-international.org/article/dynamics-content-popularity-social-media/38952