Bayerstadler, Andreas; van Dijk, Linda; Winter, Fabian
Bayesian multinomial latent variable modeling for fraud and abuse detection in health insurance. (English) Zbl 1373.62510
Insur. Math. Econ. 71, 244-252 (2016).

Summary: Healthcare fraud and abuse are a serious challenge to healthcare payers and to the entire society. This article presents a predictive model for fraud and abuse detection in health insurance based on a training dataset of manually reviewed claims. The goal of the analysis is to predict different fraud and abuse probabilities for new invoices. The prediction is based on a wide framework of fraud and abuse reports which examine the behavior of medical providers and insured members by measuring systematic deviation from usual patterns in medical claims data. We show that models which directly use the results of the reports as model covariates do not exploit the full potential in terms of predictive quality. Instead, we propose a multinomial Bayesian latent variable model which summarizes behavioral patterns in latent variables, and predicts different fraud and abuse probabilities. The estimation of model parameters is based on a Markov Chain Monte Carlo (MCMC) algorithm using Bayesian shrinkage techniques. The presented approach improves the identification of fraudulent and abusive claims compared to different benchmark approaches.

MSC:
62P05 Applications of statistics to actuarial sciences and financial mathematics
62J12 Generalized linear models (logistic models)
91B30 Risk theory, insurance (MSC2010)

Keywords:
fraud and abuse detection; health insurance; predictive model; Bayes; latent variable

Software:
Mcmcpack; MASS (R); R

Full Text: DOI

References:
Venables, W. N.; Ripley, B. D., Modern applied statistics with S, (2002), Springer
Major, J. A.; Riedinger, D. R., EFD: A hybrid knowledge/statistical-based system for the detection of fraud, J. Risk Insurance, and the Zbl 0744.62098
Fahrmeir, L.; Raach, A., A Bayesian semiparametric latent variable model for mixed responses, Psychometrika, 72, 327-346, (2007) · Zbl 1286.62104
Goldberger, A. S., Structural equation models in the social sciences, Econometrica, 979-1001, (1972)
Green, P. J., Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, Biometrika, 82, 711-732, (1995) · Zbl 0861.62023
Robert, C. P.; Casella, G., Monte Carlo statistical methods, (2004), Springer New York · Zbl 1096.62003
Venables, W. N.; Ripley, B. D., Modern applied statistics with S, (2002), Springer · Zbl 1006.62003
Viaene, S.; Derrig, R. A.; Dedene, G., Cost-sensitive learning and decision making for massachusetts PIP claim fraud data,


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.