Cost-Sensitive Learning and Decision Making for Massachusetts PIP Claim Fraud Data

Stijn Viaene¹, Richard Derrig², and Guido Dedene¹

¹ KBC Insurance Research Chair
Leuven Institute for Research on Information Systems
Department of Applied Economic Sciences
Katholieke Universiteit Leuven
Naamsestraat 69, B-3000 Leuven, Belgium
{stijn.viaene;guido.dedene}@econ.kuleuven.ac.be
² Automobile Insurers Bureau of Massachusetts & Insurance Fraud Bureau of Massachusetts
101 Arch Street, Boston, Massachusetts
richard@aib.org

Abstract. In many real-life decision making situations the default assumption of equal (mis-)classification costs underlying pattern recognition techniques is most likely violated. Consider the case of insurance claim fraud detection for which an early claim screening facility is to be built to decide upon the nature of an incoming claim as either suspicious or not. This decision typically forms the basis for routing the claim through different claims handling workflows. Claims that pass the initial (automated) screening phase are settled swiftly and routinely, involving a minimum of transaction processing costs. Claims that are flagged as suspicious need to pass a costly state verification process, involving (human) resource-intensive investigation. Here, cost-sensitive learning and decision making bring help for making cost-benefit-wise optimal decisions. In this paper we investigate the issue of cost-sensitive classification for a data set of Massachusetts closed personal injury protection (PIP) insurance claims that were previously investigated for suspicion of fraud by domain experts and for which cost information has been obtained. After a theoretical exposition on cost-sensitive learning and decision making methods, we then apply these methods to the claims data at hand to contrast the predictive performance of the documented methods for a variety of decision tree and rule learners. Standard logistic regression and (smoothed) naive Bayes are used as benchmarks.

Keywords: Personal Injury Protection (PIP) Insurance, Claim Fraud, Decision Tree, Rule Learner, Cost-Sensitive Classification.