Chapter 4
Efficient Risk Profiling Using Bayesian Networks and Particle Swarm Optimization Algorithm

Goran Klepac
Raiffeisenbank Austria d.d., Croatia

Leo Mrsic
IN2data Ltd. Data Science Company, Croatia

Robert Kopal
IN2data Ltd. Data Science Company, Croatia

ABSTRACT

Chapter introduce usage of particle swarm optimization algorithm and explained methodology, as a tool for discovering customer profiles based on previously developed Bayesian network (BN). Bayesian network usage is common known method for risk modelling although BN's are not pure statistical predictive models (like neural networks or logistic regression, for example) because their structure could also depend on expert knowledge. Bayesian network structure could be trained using algorithm but, from perspective of businesses requirements model efficiency and overall performance, it is recommended that domain expert modify Bayesian network structure using expert knowledge and experience. Chapter will also explain methodology of using particle swarm optimization algorithm as a tool for finding most riskiness profiles based on previously developed Bayesian network. Presented methodology has significant practical value in all phases of decision support in business environment (especially for complex environments).

INTRODUCTION

Bayesian networks are a method commonly used in risk modeling, but they are not a purely statistical predictive model (like, for example, neural networks or logistic regression) because their structure can also depend on expert knowledge. Bayesian network structure could be settled in algorithmic way, but from the business perspective or the perspective of model efficiency and overall performance, it is recommended that Bayesian network structure be modified by expert knowledge.

DOI: 10.4018/978-1-4666-9458-3.ch004

Expert involvement in network structure cannot be a guarantee that network will be optimal, but it is aligned with business perception of problem space. This chapter will present a proposal for additional hypothesis testing regarding the model's reliability by introducing additional tests within the network following expert involvement in network structure.

These methodologies are based on info gain measurements for each node in network which has an influence on the observed node in the Bayesian network. Using proposed methodology, expert knowledge will be challenged as well as presumed hypothesis about influences between nodes (influences between objects or business events). As a result of proposed methodology, final network should have significantly greater predictive power and greater stability and robustness.

The chapter will also present the methodology of using particle swarm optimization algorithm, as a tool for finding the most risky profiles, based on a previously developed Bayesian network. It means that particle swarm optimization algorithm will be used as a tool for finding optimal values of input variables (within developed predictive models) as referent values for maximization of probability value for some risky event. Main advantage of proposed solution is automatic determination of the most risky profiles in situation where there will be combinatory explosion caused by numerous nodes and their states within a Bayesian network model. Proposed methodology will be illustrated through a case study from insurance industry.

The case study will show situation in which insurance company wants to evaluate risk degree for each of the existing members of their portfolio. Methodology starts with Bayesian network development using automated algorithms and measuring their performance. Next step will show effects of including expert knowledge into Bayesian network structure as well as modifications made based on expert knowledge. Final step will show the effect of the proposed methodology usage, which takes into account info gain measurements for each node in network, which has an influence on observed node in the Bayesian network.

The case study will present the usage of the particle swarm optimization algorithm as a tool for finding the most risky profiles based on previously developed Bayesian network. As explained in introduction, particle swarm optimization algorithm will be used as a tool which should find optimal values of input variables (within developed predictive models) as referent values for maximization of probability value of some risky event. It means that particle swarm optimization algorithm will be used as a tool which should find optimal values of input variables within developed predictive models as referent values for maximization value of probability that customer will select/ buy some product or service. Using given results, insurance company can make profiles of the most risky insurance users, even in the situation affected by combinatory explosion, caused by numerous nodes and their states within Bayesian network model. This approach will help company to better understand a cause of riskiness of events in an environment where numerous factors and influence overlaps exist.

The basic idea is a holistic process which includes development of multinomial predictive model and usage of this model developed on historical data sample for finding risky profiles by using particle swarm optimization algorithm. This problem is not commonly expressed in case of predictive models with binomial outputs, which represents probability of buying some product or service. Bayesian networks by their definition are mostly related on multinomial outputs, not only on target variable. Reason for that lies in the fact that each connected variable within the model could be observed as an aim variable. That is the reason why those models are much more complex than, for example, models based on linear regression and binominal output.