

Auto claim fraud detection using Bayesian learning neural networks

S. Viaene^{a,b,*}, G. Dedene^{b,c}, R.A. Derrig^d

^a*Applied Economic Sciences, K. V. Leuven, Naamsestraat 69, B-3000 Leuven, Belgium*

^b*Vlerick Leuven Gent Management School, Reep1, B-9000 Gent, Belgium*

^c*Economics and Econometrics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands*

^d*Automobile Insurers Bureau of Massachusetts & Insurance Fraud Bureau of Massachusetts, 101 Arch Street, Boston MA 02110, USA*

Abstract

This article explores the explicative capabilities of neural network classifiers with automatic relevance determination weight regularization, and reports the findings from applying these networks for personal injury protection automobile insurance claim fraud detection. The automatic relevance determination objective function scheme provides us with a way to determine which inputs are most informative to the trained neural network model. An implementation of MacKay's, (1992a,b) evidence framework approach to Bayesian learning is proposed as a practical way of training such networks. The empirical evaluation is based on a data set of closed claims from accidents that occurred in Massachusetts, USA during 1993.

© 2005 Elsevier Ltd. All rights reserved.

JEL classification: C45

Keywords: Automobile insurance; Claim fraud; Neural network; Bayesian learning; Evidence framework

SIBC: IB40

1. Introduction

In recent years, the detection of fraudulent claims has blossomed into a high-priority and technology-laden problem for insurers (Viaene, 2002). Several sources speak of the increasing prevalence of insurance fraud and the sizeable proportions it has taken on (see, for example, Canadian Coalition Against Insurance Fraud, 2002; Coalition Against Insurance Fraud, 2002; Comité Européen des Assurances, 1996; 1997). September 2002, a special issue of the Journal of Risk and Insurance (Derrig, 2002) was devoted to insurance fraud topics. It scopes a significant part of previous and current technical research directions regarding insurance (claim) fraud prevention, detection and diagnosis.

More systematic electronic collection and organization of and company-wide access to coherent insurance data

have stimulated data-driven initiatives aimed at analyzing and modeling the formal relations between fraud indicator combinations and claim suspiciousness to upgrade fraud detection with (semi-)automatic, intelligible, accountable tools. Machine learning and artificial intelligence solutions are increasingly explored for the purpose of fraud prediction and diagnosis in the insurance domain. Still, all in all, little work has been published on the latter. Most of the state-of-the-art practice and methodology on fraud detection remains well-protected behind the thick walls of insurance companies. The reasons are legion.

Viaene, et al. (2002) reported on the results of a predictive performance benchmarking study. The study involved the task of learning to predict expert suspicion of personal injury protection (PIP) (no-fault) automobile insurance claim fraud. The data that was used consisted of closed real-life PIP claims from accidents that occurred in Massachusetts, USA during 1993, and that were previously investigated for suspicion of fraud by domain experts. The study contrasted several instantiations of a spectrum of state-of-the-art supervised classification techniques, that is, techniques aimed at algorithmically learning to allocate data objects, that is, input or feature vectors, to a priori defined

* Corresponding author. Tel.: +32 16 32 68 91; fax: +32 16 32 67 32.
E-mail address: stijn.viaene@econ.kuleuven.ac.be (S. Viaene).

object classes, based on a training set of data objects with known class or target labels. Among the considered techniques were neural network classifiers trained according to MacKay's (1992a) evidence framework approach to Bayesian learning. These neural networks were shown to consistently score among the best for all evaluated scenarios.

Statistical modeling techniques such as logistic regression, linear and quadratic discriminant analysis are widely used for modeling and prediction purposes. However, their predetermined functional form and restrictive (often unfounded) model assumptions limit their usefulness. The role of neural networks is to provide general and efficiently scalable parameterized nonlinear mappings between a set of input variables and a set of output variables (Bishop, 1995). Neural networks have shown to be very promising alternatives for modeling complex nonlinear relationships (see, for example, Desai et al. 1996; Lacher et al. 1995; Lee et al. 1996; Mobley et al. 2000; Piramuthu, 1999; Salchenberger et al. 1997; Sharda & Wilson, 1996). This is especially true in situations where one is confronted with a lack of domain knowledge which prevents any valid argumentation to be made concerning an appropriate model selection bias on the basis of prior domain knowledge.

Even though the modeling flexibility of neural networks makes them a very attractive and interesting alternative for pattern learning purposes, unfortunately, many practical problems still remain when implementing neural networks, such as What is the impact of the initial weight choice? How to set the weight decay parameter? How to avoid the neural network from fitting the noise in the training data? These and other issues are often dealt with in ad hoc ways. Nevertheless, they are crucial to the success of any neural network implementation. Another major objection to the use of neural networks for practical purposes remains their widely proclaimed lack of explanatory power. Neural networks are black boxes, it says. In this article Bayesian learning (Bishop, 1995; Neal, 1996) is suggested as a way to deal with these issues during neural network training in a principled, rather than an ad hoc fashion.

We set out to explore and demonstrate the explicative capabilities of neural network classifiers trained using an implementation of MacKay's (1992a) evidence framework approach to Bayesian learning for optimizing an automatic relevance determination (ARD) regularized objective function (MacKay, 1992; 1994; Neal, 1998). The ARD objective function scheme allows us to determine the relative importance of inputs to the trained model. The empirical evaluation in this article is based on the modeling work performed in the context of the baseline benchmarking study of Viaene et al. (2002).

The importance of input relevance assessment needs no underlining. It is not uncommon for domain experts to ask which inputs are relatively more important. Specifically,

Which inputs contribute most to the detection of insurance claim fraud? This is a very reasonable question. As such, methods for input selection are not only capable of improving the human understanding of the problem domain, in casu the diagnosis of insurance claim fraud, but also allow for more efficient and lower-cost solutions. In addition, penalization or elimination of (partially) redundant or irrelevant inputs may also effectively counter the curse of dimensionality (Bellman, 1961). In practice, adding inputs (even relevant ones) beyond a certain point can actually lead to a reduction in the performance of a predictive model. This is because, faced with limited data availability, as we are in practice, increasing the dimensionality of the input space will eventually lead to a situation where this space is so sparsely populated that it very poorly represents the true model in the data. This phenomenon has been termed the curse of dimensionality. The ultimate objective of input selection is, therefore, to select a minimum number of inputs required to capture the structure in the data.

This article is organized as follows. Section 2 revisits some basic theory on multilayer neural networks for classification. Section 3 elaborates on input relevance determination. The evidence framework approach to Bayesian learning for neural network classifiers is discussed in Section 4. The theoretical exposition in the first three sections is followed by an empirical evaluation. Section 5 describes the characteristics of the 1993 Massachusetts, USA PIP closed claims data that were used. Section 6 describes the setup of the empirical evaluation and reports its results. Section 7 concludes this article.

2. Neural networks for classification

Fig. 1 shows a simple three-layer neural network. It is made up of an input layer, a hidden layer and an output layer, each consisting of a number of processing units. The layers are interconnected by modifiable weights, represented by the links between the layers. A bias unit is connected to each unit other than the input units. The function of a processing unit is to accept signals along its incoming connections and (nonlinearly) transform a weighted sum of these signals, termed its activation, into a single output signal. In analogy with neurobiology, the units are sometimes called neurons. The discussion will be restricted to the use of neural networks for binary classification, where the input units represent individual components of an input vector, and a single output unit is responsible for emitting the values of the discriminant function used for classification. One then commonly opts for a multilayer neural network with one hidden layer. In principle, such a three-layer neural network can implement any continuous function from input to output, given a sufficient number of hidden units, proper nonlinearities and weights (Bishop, 1995). We start with a description of the feedforward operation of such a neural

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات