Original Article

Comparative study of modeling on claim frequency in non-life insurance

K M Sakthivel[1], C S Rajitha[2*]

[1,2]Department of Statistics, Bharathiar University, Coimbatore - 641046, Tamil Nadu, INDIA.
Email: rajitha.sugun@gmail.com

Abstract

In non-life insurance, fixing the premium for the policy is an important task. The premium of any policy mainly depends on two components, namely expected claim frequency and expected claim size. Hence the accurate and authentic estimate of the number of claim occurrences and claims size is extremely important. Different methods are available in the literature (eg; generalized linear models (GLMs), Poisson regression models etc) for estimating the claim frequency of a policy for upcoming years. But due to the heterogeneous nature of policies all these methods do not provide precise and consistent estimate of future claim frequencies. Moreover these conventional statistical methods depends mainly on some restrictive assumptions such as linearity, normality, independence among predictor variables and a pre-existing functional form relating the criterion variable and predictive variables etc. Recent studies about artificial intelligence show that an artificial neural network (ANN) is a dominant tool for prediction tasks due to some specific properties, which includes adaptive learning, nonlinearity and nonparametric nature. In this paper, we developed a procedure for estimating the future claim frequency of an insurance portfolio in non-life insurance using ANN and compare the result with GLM with suitable illustration.

Key Words: Claim Frequency; Artificial Neural Networks; Generalized Linear Model.

[*]**Address for Correspondence:**
Dr. C.S.Rajitha, Department of Statistics, Bharathiar University, Coimbatore - 641046, Tamil Nadu, INDIA.
Email: rajitha.sugun@gmail.com

Received Date: 12/04/2017 Revised Date: 11/06/2017 Accepted Date: 04/07/2017

Access this article online	
Quick Response Code:	Website: www.medpulse.in
	DOI: 02 August 2017

INTRODUCTION

One of the major crises confronted by the non-life insurance industry is the appropriate pricing of an insurance policy. With the intention of reasonably pricing insurance policies, insurers are required to estimate the average loss or pure premium in an exact manner. The pure premium of an insurance company to an insured individual is determined from estimates of two components: expected claim frequencies and expected claim size. The claim frequency for any individual stands for the number of claims that an insurer looks forward to observe in a specified year. The claim size looks up the cost related with each claim. The complexity to estimate the claim size and claim frequency in the insurance industry[8] is extensively reported in literature. In the earlier period, detecting the credibility of the insurance data sets is the main dilemma tackled by the insurance companies[22]. In general, the data sets were outsized or complex in non-life insurance. Furthermore questions take place resembling missing values and mismatched or valueless records. However the information processing job is further convincing than ever before as a result of the highly developed current IT systems. Then the dispute is to take advantage of an appropriate statistical technique to assess insurance data. Claims and risks have long been estimated as a pure algorithmic technique or a simple stochastic technique[23]. For that reason the traditional methods terminated in poor estimations. One of the traditional statistical methods for estimating the claim frequency is generalized linear models (GLMs)[14,10]. After some time modelling of both claim size and claim frequency in the presence of covariates is given by Renshaw[17]. And also Jorgensen and de Souza[9] suggested a Poisson arrival of claims and Gamma random varieties for individual costs called Tweedie model to estimate the claim frequency and claim size. Tweedie GLM has been a widely used method in actuarial studies because of its capability to model the zeros and the continuous positive outcomes at the same time[11,13,16]. But problem with the Tweedie distribution model (Tweedie GLM) is that more explanatory variables are likely to be found significant[20]. Under the normal conditions, Poisson model is conceived as the modeling representative of the frequency of claims in actuarial literature[1]. In recent times, a gradient tree-boosting algorithm[21] is proposed and apply it to Tweedie compound Poisson models for predicting pure premiums. In this study we concentrate only on the estimation of claim frequency of an insurance portfolio in general insurance.

But due to regularly increasing and altering nature in the pattern of insurance data, it is very much unstable and unsmooth. As a result all the aforementioned methods do not give accurate and consistent estimation of an insurance policy. Modern studies show that ANN can be used as an alternative to all these methods due to a number of properties, which contains adaptive learning, nonlinearity and nonparametric nature. In this paper ANN is used for modeling the insurance claim frequency and compares the efficiency of this method with GLMs. The organization of the paper is as follows: In Section 2 details about GLMs are given. In Section 3 give details about ANNs. Details of data and software used for this paper are given in section 4. Section 5 provides results. Finally in section 6 discussions and conclusions are provided.

GENERALIZED LINEAR MODELS IN NON-LIFE INSURANCE

This section provides a concise introduction to GLMs. McCullah and Nelder[10] present an absolute behaviour of the theory and application of GLMs. And also the successful applications of GLMs to actuarial problems have been analyzed by Haberman and Renshaw[6]. The GLMs is inspired in the initial occasion by the statement that the data experimented from a single-parameter exponential family of distributions. Initially we explain a number of their elementary properties in conditions of a single observation x. The log-likelihood function of a single-parameter exponential family of distributions can be written in the form

$$l = \frac{x\lambda - b(\lambda)}{\phi} + c(x,\phi) \quad (1)$$

where λ and ϕ are called the canonical and the dispersion parameters and they are presumed known. It is then clear-cut to make obvious that

$$mean = a = E(Z) = \frac{d}{d\lambda} b(\lambda) \quad (2)$$

$$\mathrm{var}(Z) = \phi \frac{d^2}{d\lambda^2} b(\lambda) = \phi\, b''(\lambda) \quad (3)$$

where $\mathrm{var}(Z)$ is the product of the quantities ϕ and $b''(\lambda)$, $b''(\lambda)$ is named variance function, be dependent on the canonical parameter and consequently on the mean. We can write this as $V(a)$. More in general a GLM is differentiated by self-determining response variables $\{Z_u; u = 1,2,...,n\}$ for which

$$E(Z_u) = a_u \quad (4)$$

$$\mathrm{var}(Z_u) = \phi V(a_u) / w_u \;, \quad (5)$$

including a variance function V, prior weights w_u and a scale parameter ϕ (> 0). Covariates to come by means of a linear predictor:

$$\eta_u = \sum_{j=1}^{p} x_{uj}\beta_j \quad (6)$$

with defined pattern (x_{uj}) and the parameters β_j which are unknown are linked to the mean response by a known differentiable monotonic link function g with

$$g(a_u) = \eta_u \quad (7)$$

$\theta(a) = \eta$, is termed the canonical link function, with the intention of the particular link function $g = \lambda$. Some Examples include the log, the, the identity, logit and the reciprocal link functions etc. The suffixes u have an inherent or obliged structure. The data consist of actualizations $\{Z_u\}$ of the independent response variables, equivalent to the construction of the units. In general, the feature of the distribution and link are predetermined, while the predictor configuration perhaps different. Maximizing the quasi-likelihood function is done for fitting the model.

NEURAL NETWORKS

Neural network (NN) model has a marvellous potential for modeling the insurance data. It does not require any preliminary information on input or output variables to complete linear and nonlinear modeling. For that reason, comparing with other methods for the estimation of the future claim frequency, ANN is more appropriate and flexible as an estimation method. And the previously presented studies have shown that the network predictions are insensible to a certain extent to diversions in the network pattern occupied[15].

Essential features of Artificial Neural Network

ANNs is machine learning method, which is considered as superior candidate for use in estimation due to its capacity to estimate unknown functions and operations. Inspired with the confidence of biological NNs, ANNs are compilation of primary processing units called artificial neurons and they are attached with each other to make a directed graph[19]. The three crucial elements of a NN are the elementary computing essentials, known as neurons or nodes, the network architecture, which depicts the association between computing units or neurons and the training algorithm, which is used to find the weights which adjust the potency of the input for the implementation of a specific responsibility. The network comprises of a set of neurons which we usually label with positive integers. The computing units are connected to each other in the sense that the output from one unit can serve as an element of the input to another. For each network connection there is a weight associated with it. Network architecture mentions the coordination of the computing units and the variety of connections permitted by the network. The most frequently used type of network in statistical applications is the multilayer feed forward network, in which the units are structured in a series of layers. Moving of information is in only one route; the units obtain information only from the upper layers of the network. The ultimate aspect of the NN is the learning or training algorithm, the method by which the values of the free parameters of the network (connection weights) are selected. The weights are generally obtained by optimizing the process output of the network on a set of training examples concerning some loss or error function. But a perceptron was treating as the original neural network. For adjusting weights in a perceptron model Rosenblatt[18] formulated the convergence procedure. Minsky and Papert[12] illustrated the weakness of perceptron convergence procedure by establishing that, perceptron models cannot categorize in the approved manner when trying to learn to identify the exclusive "or" problem. Many of the constraints possessed by two-layer networks can be defeated by multilayer networks with one or more hidden processing layers for representing complex nonlinear problems.

Multilayer Feed forward Networks

Usually two, three, or four layers of computing units consist in the NNs while it is used for statistical applications. Network receives information only from computing units in the first layer. That means the first layer serves as the input layer. The output layer or last layer develops the model solution. In the middle of these two layers exists the hidden layers and they are very important for NN to recognize the composite structure in the data. The units in the hidden layers are called hidden units as they are not able to be seen by the user. Whereas the input and output units are visible to the user. In a theoretical mode, multilayered feed forward NNs are universal estimators, and concerning its intrinsic nature, it has a remarkable ability of estimating any nonlinear mapping to any degree of accuracy[7]. They do not need a priori model to be assumed or a priori assumptions to be made on the properties of data[3]. They have been extensively employed for modeling, prediction, classification, optimization, and control purposes[2,5,4].

Primarily these multilayer networks having composite structure with hidden layers were not utilized because well-organized training algorithms were not generally accessible to permit the network to distinguish and learn the necessitated difficult nonlinear processing. It was not up to the progress of the back-propagation training algorithms[18] that it turned possible to powerfully calculate and use multilayer networks.

The Back-Propagation Algorithm

In a multilayer feed forward network, the back-propagation algorithm can be considered as a gradient search technique where the objective function is to reduce the MSE between the computed outputs of the network related to the particular set of inputs. An input pattern vector is given to the network for training the network in a consecutive behavior and carries out the computations throughout the network up to a time that an output vector is received. Then compare the computed output with the actual output for the input, and an output error is calculated. For lessening the observed prediction error the network attempts to learn by fine-tuning the weights at every distinctive neural processing unit.

DATA ANALYSIS

In this paper we estimate the expected claim frequency using ANN with the Swedish motor third party liability insurance data. It contains aggregated data on all insurance policies and claims in the year 1977. The data set consist of a data frame with 2182 observations on seven variables. The estimation of claim frequency or claim probability is based on five rating factors, the distance travelled by the vehicle, seven geographic zones, seven categories of bonus class, nine types of automobiles and the number of insured in each policy year. Frequency plot of the number of claims is given in figure 1. From the frequency plot it is observed that more number of zero claims exist in the data. So for modeling the claim frequency or number of claims, we consider both the whole data and data corresponding to the positive number of claims.

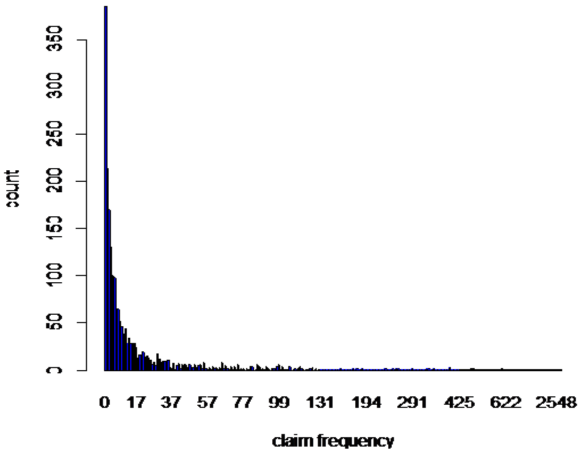

Figure 1: Frequency distribution of the number of claims

For modeling the number of claims or claim frequency we need to select the input variables and output variables. The description of input variables for modeling the number of claims is given in table. 1. Number of claims is the only one target variable

Table 1: Input variables

Sl.No	Name of the variables	Number of category
1	Distance driven by a vehicle	Five
2	Geographic zones	Seven
3	Driver claims experience	Seven
4	Types of automobile	Nine
5	Number of insured in policy years	One

Application of GLM and ANN in estimation of claim frequency

For estimating the claim frequency using GLM and ANN, initially the data set is randomly divided in to two sets, training set and testing set. For partitioning the data set we used some ratios[24]. Here we used 65:35 ratios for dividing the data set. Since the data set contains more number of zeros, the modeling of claim frequency has done by using the whole data set and the data corresponding to positive number of claims separately. Table 2 shows the data partition for all claims and table 3 shows the data partition for positive claims.

Table 2: Ratio of partition of data for all claims

Classification	Partition of data	Number of observations
Training	65%	1418
Testing	35%	764
Total	100%	2182

Table 3: Ratio of partition of data for positive claims

Classification	Partition of data	Number of observations
Training	65%	1168
Testing	35%	629
Total	100%	1797

65% of the data contained in the training set and remaining 35% of the data goes to the testing set. In GLM we used R software for estimation of claim frequencies. For estimating the claim frequency using ANN initially we train 65% of the given data using feed forward NN with back propagation algorithm. The training process provides some weights necessary to fine tune the result. The remaining 35% of data used for testing by the weights generated by the training process. Claim probabilities (claim numbers) are estimated based on the other variables such as distance travelled by the vehicle, seven geographic zones, seven categories of bonus class, nine types of automobiles and the number of insured in each policy year. Different network structures used for modeling the claim frequency of a policy is given in table 4. The outcome is estimated for all claims (Figure. 1 & Figure. 2) and for positive claims (Figure. 3 & Figure. 4). Then find the deviation between actual claim frequency and estimated claim frequency using mean square error function.

Table 4: Different structures of NN

Single hidden layer	Double hidden layer
5-8-1	5-3-2-1
5-10-1	5-4-3-1
5-20-1	5-2-1-1

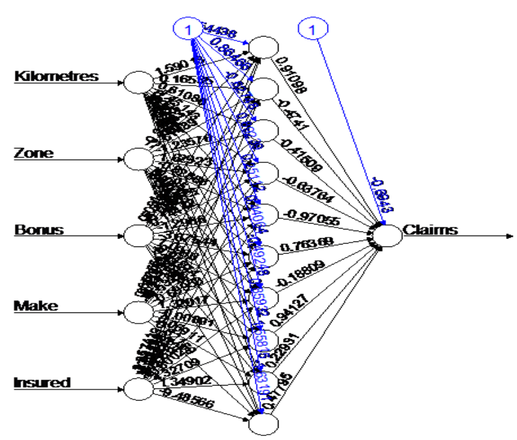

Figure 4: Structure of single hidden layer NN for all claims when h=10

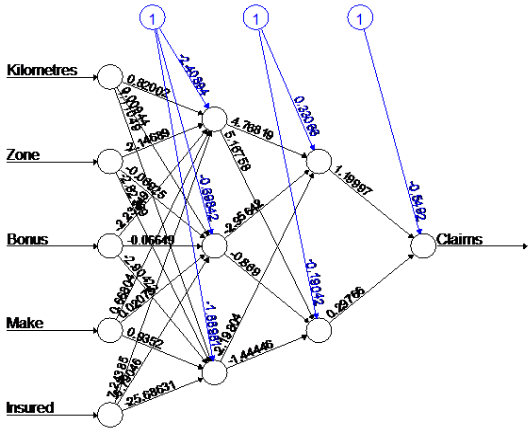

Figure 2: Structure of double hidden layer NN for all claims when h=c (3, 2)

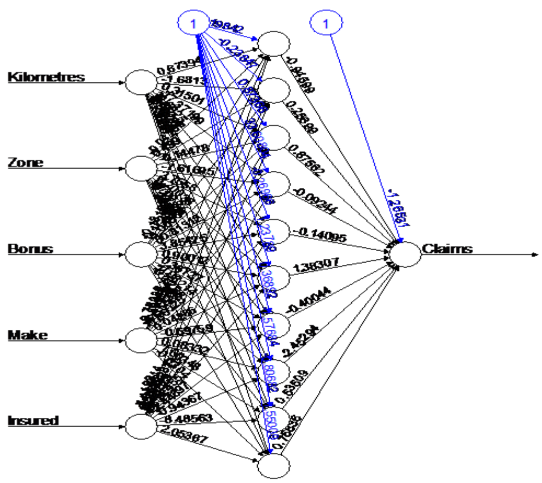

Figure 3: Structure of single hidden layer NN for positive claims when h=10

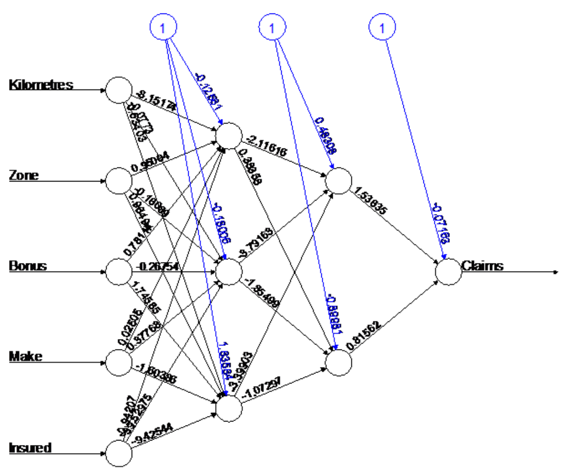

Figure 4: Structure of double hidden layer NN for positive claims when h=c (3, 2)

RESULTS

Here we compare the models using mean squared error (MSE) and also we prefer the weights of the neural network in such way that the MSE on training set is minimum. It can be defined as

$$MSE = \frac{1}{N}\sum_{j=1}^{N}(\hat{y}_j - y_j)^2 \qquad (8)$$

y_j and $\hat{y}_j$ represent the target value and network output for the jth training pattern respectively, and N is the number of training patterns.

Table 5 and Table 6 shows the standardized MSE's between the actual claim frequency and estimated claim frequency using GLM and ANN. Here the standardized mean squared error is calculated for ANN using both single hidden layer network and double hidden layer network with different number of neurons in each of the hidden layer. The estimation of claim frequency is completed by considering all claims (2182) and positive number of claims (1797) in the data. The results shows that standardized MSE between actual claim frequency and estimated claim frequency of ANN model with various hidden layers in the neural network model is smaller compared to the standardized MSE of GLM method.

Table 5: Standardised MSE using GLM

Model	MSE	
	Positive Claims	All claims
GLM	0.00105	0.00025

Table 6: Standardised MSE using ANN

Hidden Layer(h(x,y))	ANN setup	MSE(Positive claims)	MSE(all claims)
Single layered -*h*(*x*) x denotes the number of neurons in the layer	*h* = 8	0.00012	0.000129
	h =10	0.00013	0.000126
	h =20	0.000018	0.00012
Two layered-*h*(*x,y*) *x,y* denotes number of neurons in each layer	*h*=c(2,1)	0.00021	0.00024
	h=c(3,2)	0.00044	0.00017
	h=c(4,3)	0.00013	0.00016

CONCLUSION

As we know the estimation of claim frequency is substantial in insurance practice to set the premium at the beginning of the insurance contract. Though various methods have been developed for different circumstances for estimation and prediction purposes, that methods depends on some limiting assumptions such as linearity, normality, independence etc and also takes comparatively more time for estimation and prediction processes. Since ANNs are trained using the already available original data the error may be reduced and also Neural Networks depends on data and their functioning gets better with increase in sample size. We compared the ANN model with GLM, a statistical method for estimating the claim frequency. It is observed from this study that ANN is suitable for estimating the claim frequency over GLM in terms of MSE

REFERENCE

1. Antonio and E.A.Valdez,” Statistical Concepts of a Priori and a Posteriori Risk Classification in Insurance,”Advances in Statistical Analysis vol. 96(2), pp. 187-224, 2012.
2. M. Azlan Hussain, “Review of the applications of neural networks in chemical process control – simulation and online implementation,” Artificial Intelligence in Engineering, vol. 13, pp. 55–68, 1999.
3. C. M. Bishop, Neural Networks for Pattern Recognition. Oxford University, Press, Oxford, 1995.
4. B. K. Bose, “Neural network applications in power electronics and motor drives—an introduction and perspective,” IEEE Transactions on Industrial Electronics, vol. 54, pp. 14–33, 2007.
5. J. G. De Gooijer and R. J. Hyndman, “25 years of time series forecasting,” International Journal of Forecasting, vol. 22, no. 3, pp. 443–473, 2006.
6. S. Haberman, A. E. Renshaw. “Generalized linear models and actuarial science,”. Statistician, vol. 45, pp. 407-436, 1996.
7. K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” Neural Networks, vol. 2, no. 5, pp. 359–366, 1989.
8. P. de Jong and G. Z. Heller Jong, Generalized linear models for insurance data**.** Cambridge: Cambridge University Press, pp. 196, 2008.
9. B. Jorgensen and M. C de Souza, “Fitting Tweedie's compound Poisson model to insurance claims data," Scandinavian Actuarial Journal, 1994, pp. 69-93, 1994.
10. P. McCullagh and J. A Nelder, ”Generalized Linear Models,” 2nd edn. London: Chapman and Hall, 1989.
11. S. J. Mildenhall, “A systematic relationship between minimum bias and generalized linear models," In Proceedings of the Casualty Actuarial Society, vol. 86, pp. 393-487, 1999.
12. M. Minsky, and S. Papert, Perceptrons: An Introduction to Computational Geometry (Boston: MIT Press), 1969.
13. K. P. Murphy, M. J. Brockman and P. K. Lee, “Using generalized linear models to build dynamic pricing systems," In Casualty Actuarial Society Forum, Winter, pp. 107-139, 2000.
14. J. Nelder and R. Wedderburn, “Generalized Linear Models," Journal of the Royal Statistical Society. Series A (General), vol. 135, pp. 370-384, 1972.
15. Neti, Chalapathy, M. H. Schneider, and E. D. Young, , “Maximally Fault Tolerant Neural Networks,” IEEE Transactions on Neural Networks, vol. 30(1), pp. 12-23, 1992.
16. G. W. Peters, P. V. Shevchenko and M. V. Wuthrich, “Model uncertainty in claims reserving within Tweedie's compound Poisson models," ASTIN Bulletin, vol. 39(1), pp. 1-33, 2009.
17. A. E. Renshaw, “Modelling the claims process in the presence of covariates," ASTIN Bulletin, vol. 24, pp. 265-285, 1994.
18. F. Rosenblatt, “Two Theorems of Statistical Separability in the Perceptron,” Proceedings of a Symposium on the Mechanization of Thought Processes (London: Her Majesty's Stationery Office), pp. 421-456, 1959.
19. S. S. Haykin. Neural networks and learning machines. Pearson, Upper Saddle River, N.J, 2009.
20. G. Smyth and B. Jorgensen, “Fitting Tweedie's compound Poisson model to insurance claims data: Dispersion modelling," ASTIN Bulletin, vol. 32, pp. 143-157, 2002.
21. Yi Yang, Wei Qian and Hui Zou, “Insurance Premium Prediction via Gradient Tree-Boosted Tweedie Compound Poisson Models,” Journal of Business and Economic Statistics, vol. 34(3), pp. 1-45, 2016.
22. H. I. Weisberg, and T. J. Tomberlin, “A Statistical Perspective on Actuarial Methods for Estimating Pure Premiums from Cross-Classified Data,” The Journal of Risk and Insurance, Vol. 49, (4), pp. 539-563, 1982.
23. Wüthrich, V. Mario and Merz, Michael. Stochastic claims reserving methods in insurance. West Sussex: John Wiley & Sons, Ltd, 2008.
24. G. Zhang, B. E. Patuwo, M.Y. Hu, “Forecasting with artificial neural networks: The state of the art,” International Journal of Forecasting, Vol. 14, pp. 35–62, 1998.Wüthrich, V. Mario and Merz, Michael. Stochastic claims

Source of Support: None Declared
Conflict of Interest: None Declared