Estimation of Products Final Price Using Bayesian Analysis Generalized Poisson Model and Artificial Neural Networks

Marjan Niyati[a,*] , Amir Masud Eftekhari Moghadam[a,*]

[a] *Department Of Computer Engineering And Information Technology Qazvin Azad University Of Technology ,IRAN*

Received 3 Nov., 2008; revised 20 Dec., 2008; accepted 12 Feb. 2009

Abstract

Estimating the final price of products is of great importance. For manufacturing companies proposing a final price is only possible after the design process over. These companies propose an approximate initial price of the required products to the customers for which some of time and money is required. Here using the existing data of already designed transformers and utilizing the bayesian analysis of generalize poisson models and artificial neural networks, a shortcut method for estimating the material and final price of transformers is established. The proposed method being quite precise and fast, without any cost.

Keywords: Bayesian analysis, Artificial neural network;

1. Introduction

The progresses made in the design and manufacturing of products are the most important subjects encountered by industrial companies. The initial estimation for special products undergoes a long and time consuming process. In many cases having an approximate price for products in the first stages of a project is very effective. In this paper a method for price estimation based on a hybrid approach of artificial neural network and bayesian analysis of generalized Poisson is proposed to optimize the network weights using the existing data of a transformer manufacturing company. First, the effective parameters on products prices are examined, then we will have a short discussion on artificial neural networks. The structure and sets of data used for training and validating the network are examined. Finally the produced results of the neural network estimation will be trained which shows the accuracy rate in price estimation and raw material [9]. Bayesian probability here means the concept of probability used in the bayesian decision theory [8].
Bayesian analysis is often hard [1,3].A simple implementation that numerically assesses the posterior is incases where the closed form of the bayesian posterior is hard to formulate, exploiting a discrete of the prior. Representation is reminiscent of importance sampling, in that we simulate the discrete points from the prior, and then weight by the ratio of the posterior to the prior, that is, by the likelihood. We have used the bayesian analysis1 as a reliable method for analyzing GLM when there are only few observations. Using regression techniques for analyzing GLM with a small sample often leads to unreliable results, including large confidence ellipsoids and high bias. Furthermore, this method allows the estimation of the model coefficients even when the number of observations is smaller than the number of coefficients [1].

2. Effective parameters on product prices

Various parameters could be effective on the final price of the products. The product of raw material price as well as labor and equipment used in manufacturing, are the most important and effective parameters of the final price. Therefore the final price could be achieved using relation [1].

*Corresponding author. Tel.: +09123417847; e-mail: m.niyati@iran-transfo.com,eftekhari@qazviniav.ac.ir

$Gp^2 = [Rmp^3] + [Laec^4]$ (1)

The prices of raw material, labor and equipment are achievable using relations (2) and (3):

$Cmp^5 = [Cmw^6] \times [Upocm^7]$ (2)

$Tcolae^8 = [Tnftac^9] \times [Coloeph^{10}]$ (3)

The raw material used in manufacturing transformers, include copper, core steel, industrial oil and insulating material. The time needed or person-hour spent on the production of transformers are also calculated by breaking down the various production processes as well as identifying the effective parameters on the time spent on each process and the simulation of this relation and its final integration . In this article, the total weight of iron wave, oil and copper are considered as the effective parameters on the final price and the production period is defined as the labor and equipment cost [9].

3. Artificial neural networks

Artificial neural network (ANN), also called simulated neural network (SNN) or commonly just neural network (NN), is an interconnected group of artificial neurons that uses a computational or mathematical model for information processing based on a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network. An ANN is composed of many artificial neurons that are linked together according to specific network architecture [5,4]. The objective of the neural network is to transform the inputs into meaningful outputs [5-10]. BPN is a common ANN architecture model whose architecture is the multilayer percept (MLP) [1].The BPN uses the idea of "the gradient steepest descent method" to minimize the errors between actual and predicted output functions. An increased number of hidden layers and the transformation function of the smoothing differential can allow the network to apply the gradient steepest descent method to correct the network weights formula. Therefore, if there is enough hidden layer or hidden neurons, the linear threshold curve can approach any function. The learning procedures of a BPN contain initialization, forwarding and reverse processes. The initial weights and bias are generated randomly to calculate the net input and output of each unit among the hidden and output layers while the deviation is processing the entry of forward propagation. The errors feed back to the network through the updated weights and bias of back-propagation in order to approximate to the minimal error. Before applying the BPN to solve problems, we must also set the network architecture and its parameters of BPN in advance, such as hidden layer, learning rate, momentum term, number of hidden neurons and learning cycle. The network architecture and its parameters must be determined carefully in order to avoid constructing a worse network model which may increase the computational cost significantly and produce worse result. In most pattern classification problems, given a large set of potential features, it is usually necessary to find a small subset with which to classify. The information without any feature selection might be redundant or nosey so that this can deteriorate or reduce the efficiency of classification. The main benefits of feature selection are as follow:

1. Reducing the computational cost and storage requirements.
2. Dealing with the degradation of the classification efficiency due.
3. R educing training and prediction time
4. Facilitating data understanding and data visualization [5].

3.1. BPN

1. Learning rate and momentum term: Too high a learning rate will cause the architecture to oscillate and be hard to converge; too low a learning rate will cause slow convergence and may fall into local optimization. Momentum term is the parameter that lets the architecture converge relatively fast. Too small a momentum term does not have an obvious effect and cannot increase the classification accuracy rate; too big a momentum term can excessively affect learning effect and cause extreme modification**.**

2. Number of hidden neurons: When there are a few numbers of hidden neurons, it causes bigger errors. Increasing the number of hidden neurons can alleviate this situation but will simultaneously affect the speeds of convergence and computing with almost no help in reducing errors after exceeding a certain degree of node numbers. In general, the number of hidden neurons is better to be small when noisy data are too large; on the other

hand, the number of hidden neurons is better to be large when the complexity of the problem is high. However, ultimately, adjustments depend on the characteristics of the data. An often suggested number is half of the features or

[2] Good price
[3] Raw material price
[4] Labor and equipment cost
[5] Consumed material price
[6] Consumed material weight
[7] Unit price of consumed material
[8] The cost of labor and equipment
[9] Time needed for technology and construction
[10] Cost of labor or equipment per hour

the square root of the number of input features times the number of output features.

3. Learning cycle: Too high a learning cycle can result in over-fitting; too low learning cycle can cause not enough training and result in worse classification accuracy rate of

Test data. In general, cycle ends when the Mean Square Error (MSE) or Root Mean-Square Error (RMSE) of training data is smaller than a certain value or reaches a pre-determined number. Therefore, rule of thumb or "try and error" methods are used to determine the network architecture and parameters. After many experiments, sometimes users may probably find out an accepted one. However, these methods are hardly to obtain a better network architecture and parameters. As a result, many people have tried to apply three types of approaches to avoid them.[8].

4. Poisson distribution

Extra-Poisson variation or over-dispersion, relative to a Poisson model. This shortcoming of the Poisson distribution is also a concern in the context of the Poisson regression, when the mean of the response variable is affected by a number of explanatory variables. For this reason, numerous authors have proposed tests for detecting over-dispersion in Poisson models. One version of the GPD has a probability function given by [5]:

$$P(Y=y \mid \theta, \lambda) = \theta\,(\theta + y\,\lambda)^{y-1}\,(y!)^{-1} \exp(-\theta - y\,\lambda) \qquad (4)$$

Where $\theta > 0$ and $0 \leq \lambda <$, 1for those values of y on the non-negative integers, and zero else where. Then $\lambda = 0$ this distribution reduces to the standard Poisson. It is well-known that the GPD has a mean of $\theta\,(1-\lambda)^{-1}$ and a variance of $\theta\,(1-\lambda)^{-3}$, and so this distribution may be suitable when count data is observed with a sample variance considerably larger than the sample mean. Often, the random count data are affected by a number of explanatory regression variables. For instance, this is the case when insurance policies are grouped according to the different levels of various risk factors, and we are model the observed number of claims by class. In this case it is easy to define a GPR model based upon the GPD. For a sample of size n, let the *i*'th response variable be denoted Y_i and let x_i denote the associated p x 1 vector of explanatory variables. With the prior bayesian practitioner's for the model parameters using bayesian theorem, the result will be the posterior distribution. In general, unless the size of the data set is very small and the prior density also has a very convenient form, the resulting posterior will not have a particularly tractable form for analytical analysis.[9]. The covariate vector x_i, lets the distribution of Y i to be that of the GPD with probability function (1) with the mean of $E(Y_{i \mid Xi}; \beta, \lambda) = \mu(Xi; \beta) = \mu i > 0$ and with the secondary parameter ; λ. Our assumption is that $\mu(Xi : \beta)$ is a known function of x_i and an associated p x 1 vector β of regression parameters. Taking into account that the mean of the GPD is given by $\mu i = \theta(1-\lambda)-1 = \theta\varphi$, the corresponding GPR model for the response variable Y_i may be written as:

$$P(Y_i = y_i \mid xi: \beta, \lambda)$$

$$= \mu_i\,{}_{[}\mu_i + (\varphi - 1)\,y_i^{\;yi-1}\,\varphi^{-y}{}_i(y_i\,!)^{-1} \exp\{-[\mu_i + ((\varphi - 1)\,y_{i\,]}/\varphi\} \qquad (5)$$

Where $\mu_i > 0$ and $0 \leq \lambda < 1$, for those values of y_i, on the non-negative integers. We previously noted that the parameter $\varphi = (1-\lambda)-1$ represents the square root of the index of dispersion, and that the variance of Y_i is

$$Var(Y_i \mid x_i; \beta, \lambda) = \varphi^2\,\mu_i > 0 \qquad (6)$$

When we are faced with a random sample that has been generated according to either the GPD or the GPR model, it is evident that the likelihood function will be formed as a product of form (1) or (2) respectively. If we simply combine this likelihood with the bayesian practitioner's prior for the model parameters using bayesian' theorem, then the result will be the posterior distribution. In general, unless the size of the data set is very small and the prior density also has a very convenient form, the resulting posterior will not have a particularly tractable form for analytical analysis. Count data is observed with a sample variance considerably larger than the sample mean. Often, the random count data are affected by a number of explanatory regression variables. For instance, this is the case when insurance policies are grouped according to the different levels of various risk factors, so the observed number of claims is modeled by class. In this case it is easy to define a GPR model based upon the GPD. For a sample of size n, let the i th response variable be denoted Y_i and let x_i denote the associated p x 1 vector of explanatory variables. [9].

5. Implementation of artificial neural network in order to estimate the product prices

Implementation of artificial neural network for transformer estimation was done using the MATLAB 7.1 software. The main steps in this implementation include:

- Training and test data
- Preprocessing of data neural network structure
- Validation of neural structure

5.1. Required data for training and test

It is essential to have a proper collection of data, for training and validation of neural networks .In the present problem, the inputs and outputs illustrated in figure1 have been considered for the neural network .The necessary data have been gathered from IRAN-TRANSFO commercial company. From the information of 2006, 57 couples of input-output where used for training and the 57 other couples for the validation of the network.

5.2. Data preprocessing

Before training, pre processing for data must be specified so that the networks training could be done in a more efficient way. For this, the input and output data have been presented in two separate matrixes. The input matrix which is used for network training , is a 57×7 matrix and the out put matrix for training is a 57×5 one in which numbers at rows are divided to the biggest number to be normalized. The same normalization is applied in the validation process as well.

5.3. The structure of neural network

The neural network structure defines the neurons, organization and their connection as well as functions used in each neuron. In this research we used a feedback 3 layer networks which consists of the input layer, a hidden layer and an output layer. The functions and number of these layers are respectively 5, 3, 7 neurons.

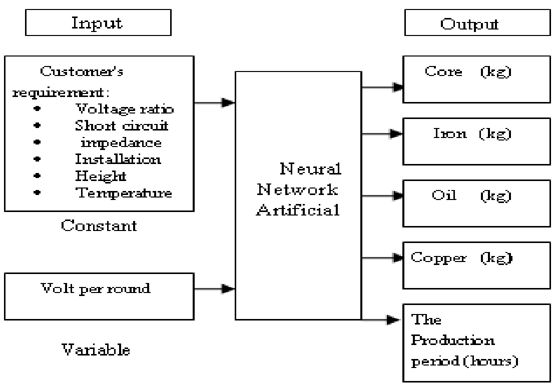

Fig.1. The plot of neural network with input and output parameters.

6. Experimental results

Applying the bayesian in poisson distribution which is adopted from GLM we come to the following results. The number of vectors to display the disparity are NN=10000 and the value of Betameans is defined as Betameans = [-0.7 0.5 0.7 0.7] and the variance is considered as an interval of Betamens = [1.5 3 1.5 1.5]. The plot of Bayesian posterior interval and regular Regression Results is illustrated in figure2.

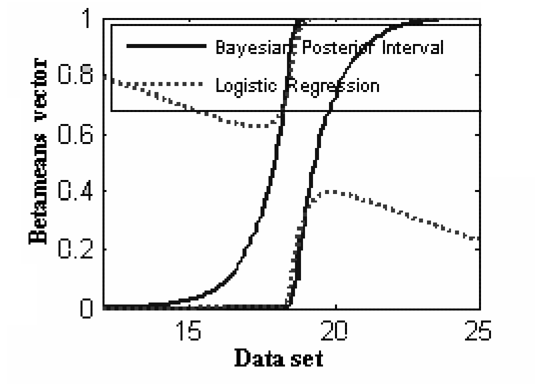

Fig.2. Bayesian Posterior interval

6.1. Training and validation of the neural network

In the present study, the neural network used, is trained via the momentum term and back-propagation laws. The previously explained data in the last section have been used in the training of the neural network. The training rate, momentum term and the final error have been considered to be 0.005, 0.5 and 10^4 .The discussed network has been trained after passing an approximate 480000 repeats in the expected errors. The plot of changes in network error while learning is illustrated in figure 3.

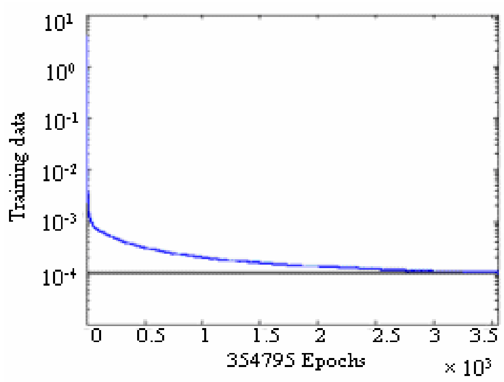

Fig.3. Changes in network error while learning

After network training, the set of relevant validation data is applied to the network and the error percentage per each out put is shown in figure 5, 6, 7,8, 9.

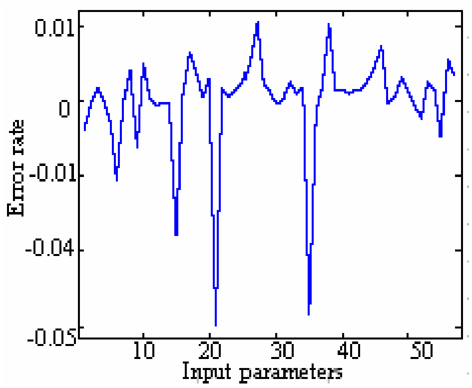

Fig.4. Errors of the output neural network in estimating the total core weights.

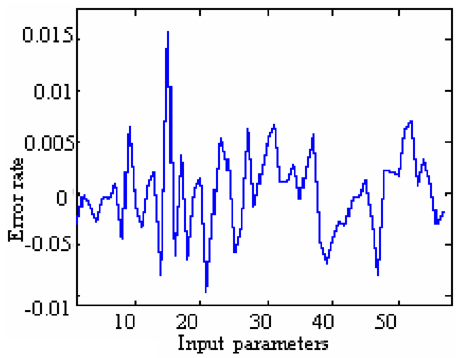

Fig.5. Error of the output neural network in estimating the iron Weight

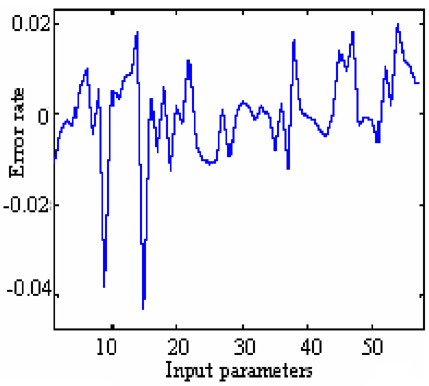

Fig.6. Error of the neural network in estimating the copper weight output

Through studying these figures, it is obvious that the maximum neural network error in estimating of the outputs of the previously trained data, is a maximum of 5.951% which is related to the estimation of the relative output of the total weight of the core, there for it is expected that the taught neural network could give better estimation in response to the previously taught data.

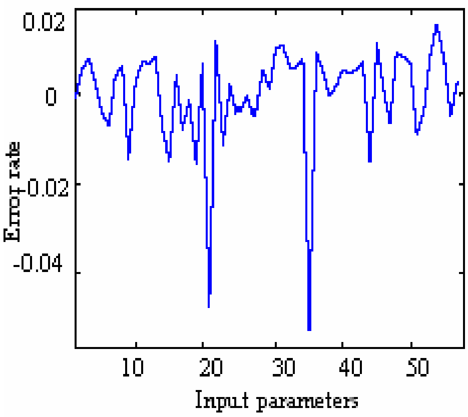

Fig.7. Error of the neural network in estimating the oil weight output

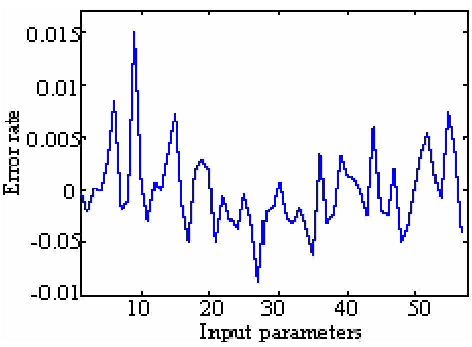

Fig.8. Error of the neural network in estimating the production period

7. Conclusion

In this article a method for estimating transformer price is presented. This article is based on the artificial neural network and Bayesian analysis on the basis of poisson propagation, due to the requirements of manufacturing companies to propose an approximate initial price without spending any extra cost or time to customers.

Through appropriate network training and the validation of the data which did not participate in the training process, the accuracy of estimation of raw materials estimation used for a transformer and the final price is proved. The presented method could be an effective way in order to estimate the final price, very fast and cost less.

References

[1] H. Adoror, D. MSteinberg, Sequential Experimental Designs for Generalized Linear Models, Journal of the American Statistical Association, vol. 103, 288-299, 2008.

[2] D.Allan., Power Transformers-The second century, Power Engineering Journal,vol. 67, 25-31.2005.

[3] J. Anneting, A.D. Souza, S. Vijayakumar, S. Schaal, A Bayesian Approach to Empirical Local Linearization for Robotics, Proc.International Computer Eng.Conf.,IEEE, Vic., 20Aug., 2008.

[4] S. Chen, S. Lin, T. Tseng ,H.Lin, Optimization of Back-Propagation Network Using Simulated Annealing Approach,Proc.30 th Research Simulation and Computation. Conf., IEEE,Vic., 19-23Apr.,2006.

[5] J. Han, M. Kamber, Data Mining: Concepts and Techniques, , in:M. kaufmann (Ed), Computer and Technoligy,vol. 3, Sanfrancisco, 2003,Ch. 3.

[6] H. Hosseini, H. Seyfi, The Optimized Design Transformer with the Aid of Genetic Algorithm, Journal of Technical Colledge of Tarbiyat Modares University, vol. 1,77-84,2002.

[7] P. Maher, Bayesian Probability, Journal of Illinois University at Urbana-Champaign, vol. 12, 65-76, 2007.

[8] A. Matsunaga, K. Ogawa, Scatter Correction in Multinuclide Data Acquisition by Means of A Neural Network,.Proc.5th Nuclear Science Symposium.Conf.,Acm ,Vic.,20 Feb., 1999.

[9] D. Scollnik, The Bayesian Analysis of Generalized Poisson Models for Claim Frequency Data Utilising Markov Chain Monte Carlo Methods, Proc.29th Actuarial Research Simulation and Computation. Conf., Acm, Vic.,20-28Feb, 1994.