Using Artificial Neural Networks to Establish a Customer-cancellation Prediction Model

Abstract. In the past, judgments concerning customer cancellations relied primarily on managers’ experience. Prediction errors can cause surpluses or insufficient service capacity. Data mining technology can improve prediction and judgment accuracy. This study applies back propagation neural networks and general regression neural networks to establish a customer-cancellation prediction model. The empirical results showed that both prediction models possessed good predictive abilities and can aid in service capacity scheduling.

Introduction

To reduce needless customer waiting and to improve service quality, businesses frequently adopt a reservation system to schedule times for customer service. However, reservations entail cancellations. Accurately predicting the number of customer cancellations is a crucial issue [1–4]. Numerous businesses primarily base customer cancellation estimates on trial-and-error, which can lead to service overcapacity and wastage of resources, or to insufficient service capacity and customer complaints [5–7].

An analysis of fundamental customer attributes and consumer records can aid customer cancellation predictions and judgments [8, 9], but extensive customer data may not be useful. Data that is extensive and disorganized can be processed efficiently and accurately using data mining technology to determine the implicit, potentially useful information [10]. Data mining is divided into classification, estimation, prediction, homogeneous grouping or association rules, clustering, description, and visualization [11–13]. The characteristics of the classification are defined according to existing information types. The classification model is established through machine learning. When the model encounters new information, it can determine the category in which this information should be placed [14, 15].

Customer cancellation is a typical classification problem and machine learning technology can be employed to increase managers’ accuracy regarding judging whether customers will cancel. Therefore, this study uses back propagation neural networks (BPN) and general regression neural networks (GRNN) to construct customer cancellation prediction models, assisting managers in making customer cancellation judgments and improving their prediction accuracy to avoid unnecessary loss regarding costs [16].

Artificial Neural Networks

Human brains contain countless neurons, and each type of neuron governs a different perception, reasoning, memory, or learning function. An artificial neural network (ANN) is an information processing system that mimics the brain’s neural organization and operations [17]. An ANN is constructed using numerous artificial neurons, and the output of every processing element becomes the input of the following processing element. The relationship equation for artificial neuron input and output can be expressed with Eq. (1):

\[ Y_j = f \left( \sum W_{ij} X_i - \theta_j \right) \]

where:
- \( Y_j \) - output signal, \( f \) - transfer function, \( W_{ij} \) - synapse strength, also known as link weighting, \( X_i \) - input signal, \( \theta_j \) - bias.

A typical ANN structure is divided into an input layer, hidden layer, and output layer [18] (Figure 1):

- Input layer: represents the ANN’s input variables.
- Hidden layer: represents the interaction influences among the neurons. An ANN may have more than one hidden layer or it could have no hidden layers.
- Output layer: represents the ANN’s output variables.

Fig. 1. ANN conceptual diagram

These can be divided into four types according to the ANN learning method: supervised learning, unsupervised learning, associative learning, and optimization application [19].

1. Supervised Learning Networks

Supervised learning networks take training examples from problems, including input and output values. First, the network must be trained, and the network weight is adjusted based on the discrepancy between the expected output value and actual calculated value.

2. Unsupervised Learning Network

Unsupervised learning networks contain input data and do not include expected output data. Therefore, there is no minimum error requirement. The training method is based on the input data. Internal classification rules for learning examples do not require external supervision to automatically adjust the network weights. This enables the
network to independently classify identical or similar output vectors for use in sample identification.

3. Associate Learning Network
This type of network learns internal associative memory rules from various examples and applies them to new cases and is typically used for data acquisition and noise filtering.

4. Optimization Application Network
This type of network sets the variable values according to the problem. Under the condition of satisfying the design constraints, it can achieve an optimized design objective.

Back Propagation Neural Network
BPN is the most representative and widely used ANN model and extends from the perceptron. Because of the lack of a hidden layer design, the perceptron’s learning ability is limited. In 1974, Werbos [20] first proposed hidden layer learning algorithms. In 1985, Parker re-proposed this idea [21] and BPN began to receive attention. A BPN is a supervised learning network that is widely used for prediction, classification, and diagnostics. It operates by first obtaining a problem’s training sample and actual output value and uses the gradient steepest descent method to repeatedly adjust and correct the weight of each neuron, minimizing the error between the output values and true values [22-24].

General Regression Neural Networks
GRNNs were proposed by Specht [25] and evolved from probabilistic neural networks (PNN). They can analyze classifications, predictions, or control problems and possess excellent processing power or ability for linear or non-linear regression problems. The GRNN learning process differs from that of the BPN as described below [26]:
- The learning process is not related to the recall process.
- It does not use the discrepancy between the inferred output and training sample target values to correct the network link weightings.
- Utilizes one-pass learning.
- The objective of the learning process is to determine the optimal smoothing parameter.
- Network neuron count and training samples are related.

Methods
This study adopted BPN and GRNN to construct a customer cancellation prediction model and used 1,400 customer reservation data taken from a Western restaurant chain in Taiwan in 2011. Of these reservations, 251 were cancelled reservations (17.93%). There were 12 customer attribute and consumer record variables: year, month, day, whether it was a holiday, gender, age, income, education level, marital status, place of residence, cancellation record, and cumulative number of cancellations. This study used 70% of the data for the training data and 30% for the test data.

Parameter Settings
The BPN and GRNN network structures all had one input layer, one hidden layer, and one output layer. There were 12 input layer neurons and one neuron in the output layer (the target values were 0 or 1, 0 signified fulfilled appointments or reservations, 1 signified cancellation). The BPN uses the transfer function tanh. It employs a search method to determine the hidden layer neuron quantity and adopts the judgment criteria of the least error for output results and actual data, as well as the highest correlation. We used a decreasing search method to set the GRNN smoothing parameter $\sigma$ [27]. This study set the starting smoothing parameter coefficient as 5, the split-half coefficient as 0.5, and the optimized smoothing parameter as 0.078125 (Table 1).

Table 1. Optimized smoothing parameter search results

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>0.5</th>
<th>2.5</th>
<th>1.25</th>
<th>0.625</th>
<th>0.3125</th>
<th>0.15625</th>
<th>0.078125</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.100227</td>
<td>0.100185</td>
<td>0.100031</td>
<td>0.099563</td>
<td>0.099393</td>
<td>0.098697</td>
<td>0.098601</td>
</tr>
</tbody>
</table>

Results and Analysis
This study analyzed the prediction model’s performance and identification abilities according to specificity, sensitivity, and receiver operating characteristics (ROC). Assuming $n_1$ actual cancellations (Type A) and $n_1$ fulfilled appointments (Type B), the prediction model’s correct predictions were divided into $n_{11}$ (Type A) and $n_{12}$ (Type B), and prediction model errors were divided into $n_{21}$ (Type A) and $n_{22}$ (Type B). The sensitivity was calculated according to Eq. 2, and the specificity was calculated according to Eq. 3.

\[
\text{Sensitivity} = \frac{n_{11}}{n_1} \quad \text{Eq. 2} \\
\text{Specificity} = \frac{n_{22}}{n_2} \quad \text{Eq. 3}
\]

The ROC curve was drawn using specificity and sensitivity. The greatest area under the curve was 1. Larger areas under the ROC curve represent the prediction model’s ability to correctly determine the data type, whereas smaller areas represent lower identification ability.

When the BPN and GRNN prediction models completed their learning, we used 420 sets of data to test the models’ accuracies. For the test results, see Tables 3 and 4. The ROC is shown in Figure 2, which indicated that BPN had greater classification accuracy results than did GRNN.

Table 3. BPN test results

<table>
<thead>
<tr>
<th>Prediction Value 1</th>
<th>Actual Value 1</th>
<th>Actual Value 0</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction Value 1</td>
<td>56</td>
<td>88</td>
<td>144</td>
</tr>
<tr>
<td>Prediction Value 0</td>
<td>14</td>
<td>262</td>
<td>276</td>
</tr>
<tr>
<td>Total</td>
<td>70</td>
<td>350</td>
<td>420</td>
</tr>
</tbody>
</table>

1: signified cancellation
0: signified fulfilled appointments

Table 4. GRNN test results

<table>
<thead>
<tr>
<th>Prediction Value 1</th>
<th>Actual Value 1</th>
<th>Actual Value 0</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction Value 1</td>
<td>61</td>
<td>107</td>
<td>168</td>
</tr>
<tr>
<td>Prediction Value 0</td>
<td>9</td>
<td>243</td>
<td>252</td>
</tr>
<tr>
<td>Total</td>
<td>70</td>
<td>350</td>
<td>420</td>
</tr>
</tbody>
</table>

1: signified cancellation
0: signified fulfilled appointments

Table 5 indicates the specificity, sensitivity, and the square under the ROC curve of the BPN and GRNN prediction models. Both models produced acceptable results for classification. Their pros and cons are as follows:

1. Specificity: GRNN > BPN;
2. Sensitivity: BPN > GRNN;
3. Square under curve: BPN > GRNN.
Table 5. Test results of both ANN models

<table>
<thead>
<tr>
<th></th>
<th>Specificity</th>
<th>Sensitivity</th>
<th>Square under curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPN</td>
<td>80.00%</td>
<td>74.86%</td>
<td>80.87%</td>
</tr>
<tr>
<td>GRNN</td>
<td>87.14%</td>
<td>69.43%</td>
<td>75.34%</td>
</tr>
</tbody>
</table>

Conclusions

To reduce waiting time for customer service, businesses frequently adopt a reservation system to schedule times for relevant services. However, reservations entail cancellations. Businesses frequently enable the number of customer reservations to exceed their service capacity, which can lead to overbooking, to prevent service overcapacity. Various businesses also rely solely on managers’ experience to estimate the number of customer cancellations, and the resulting prediction errors can result in surpluses or insufficient service capacity. This study applied BPNs and GRNNs to establish customer cancellation prediction models. The empirical results showed that both prediction models possessed excellent cancellation predictive abilities, can assist managers in judging whether customers will cancel reservations, and can aid in dynamic service capacity scheduling.

REFERENCES


Authors: prof. Han-Chen Huang, Department of Leisure Management, Yu Da University, No. 168, Hsueh-fu Rd., Tan-wen Village, Chao-chiao Township, Miaoli County, 36143 Taiwan, R.O.C., E-mail: tchhuang@ydu.edu.tw; Corresponding author: prof. Allen Y. Chang, Department of Computer Science and Information Engineering, Chinese Culture University, No. 55 Hwa-Kang Rd., Yang-Ming-Shan, Taipei, Taiwan, 11114, R.O.C., E-mail: zyh3@faculty.pccu.edu.tw; prof. Chih-Chung Ho, Graduate Institute of Earth Science, Chinese Culture University, No. 55 Hwa-Kang Rd., Yang-Ming-Shan, Taipei, Taiwan, 11114, R.O.C., E-mail: hzz6@ulive.pccu.edu.tw.