Expanding Horizons
Business, Management and
Technology for Better Society
20–22 May 2020
Online Conference

make learn TIIM

Management,
Knowledge and Learning
International Conference 2020
Technology, Innovation
and Industrial Management

Analysis of E-learning Customer Data using Data Mining Techniques

Kuo-Ping Lin
Department of Industrial Engineering and Enterprise Information, Tunghai University, Taiwan
kplin@thu.edu.tw

Yu-Ming Lu
Department of Information Management, Lunghwa University of Science and Technology, Taiwan
yuminglu@ms2.hinet.net

Chih-Hung Jen
Department of Information Management, Lunghwa University of Science and Technology, Taiwan
f7815@mail.lhu.edu.tw

Ming-Jyun Chiang
Department of Industrial Engineering and Enterprise Information, Tunghai University, Taiwan
G08330015@thu.edu.tw

Abstract

The purpose of this study is that provide analysis of e-learning customer data by using machine learning methods include decision tree, Deep belief network (DBN), and support vector machine (SVM). E-learning marketing need to precisely understand their customer in e-learning industry. Deep belief network (DBN) models have been successfully employed to classify problem. This study uses a three-layer deep network of restricted Boltzmann machines (RBMs) to capture the feature of input space of customer data, and after pre training of RBMs using their energy functions, gradient descent training. The customer data of e-learning courses was collected and examined to determine the feasibility of the decision tree, DBN and SVM. This study uses the actual database to select customer's data include "sex", "birth month", "public/private university", "home postal code", and decision variable "classes of study". These customer's datasets are examined through decision trees, support vector machines, and Deep Belief Network Classifier, which provides rules and classifier training results for digital marketing systems. This study can help exploring the relationship of courses, and promote the ability of information for e-learning enterprise. The results show that (1) male students almost selected engineering courses, and (2) female students almost selected business courses. Mainly, except those who live south of Changhua and were born after March or students who were born after September and are "non-Taipei". (3) Students from public or private universities will not affect the students' willingness to study e-learning courses.

Keywords: deep belief network, support vector machines, decision trees, de-learning course

INTRODUCTION

Everyone wants to learn different kinds of things, and they can select different courses on e-learning platform. E-learning customers can choose via internet according to the course name and course content. For digital learning platform companies, they want customers to learn more courses via learning platform, but they do not want customers to choose courses that are not interesting or do not need customers at this stage. Therefore, this study adopt machine learning technique which includes decision tree, deepand support vector mechine to evaluate E-learning customers based on customer data. Some studies have used the machine learning to analyze customer data in various industries such as bank customers (Chen, 2020), hotel customers (Dursun and Caber, 2016),…, etc. The rest of this paper is organized as follows: The machine learning methodologies are introduced in section 2; in section 3, numerical examples of e-learning customer datasets are utilized to demonstrate the performance of different methods; finally, conclusions are made in section 4.

Methodology

Decision tree-Quinlan (1986) proposed the Iterative Dichotomiser 3 **(**ID3) algorithm. This algorithm is based on: "Smaller decision trees are better than large decision trees", that is, simple theory. Therefore, the main goal of ID3 is to choose the appropriate Attributes are used as nodes. The index used when selecting attributes is based on Information Gain. Entropy is used to measure the consistency of data. When the entropy value is greater, the data is more cluttered. Assuming the target has attributes with N different values, the entropy (*S*) of the classification relative to *N* states is defined as

$$Entropy(S) = \sum_{i=1}^{N} -p_i \log_2 p_i \tag{1}$$

Where p_i is the probability of occurrence of each state.

The ID3 algorithm will use the information profit to measure the ability of each attribute to classify the data. The information gain Gain (*S*, *E*) is used to evaluate the *V* data of the attribute E in the S set, which is defined as:

$$Gain(S,E) = Entropy(S) - \sum_{j=1}^{v} \frac{|S_j|}{S} Entropy(S_j) \tag{2}$$

This *Gain* value evaluates the ability to classify data by the degree of internal data clutter. The smaller the *Gain* value means the more cluttered the data, and classification ability is worse. Conversely, the larger the *Gain* value measns the less cluttered the data, and classification ability is better. The ID3 algorithm will select the attribute with the most information as the classification attribute.

Support vector machine– Originally, SVM was designed for two-class classification. Based on the process of determining the separate boundary and the maximum distance to the closest points, SVM derives a class choice, called support vectors (SVs), for the training data set. SVM can avoid a potential misclassification in the testing data by minimizing structural risk rather than empirical risk. Therefore, the SVM classifier demonstrates better generalization performance than that of other traditional classifiers. First, we give a training data set $D = \{x_i, Y_i\}_{i=1}^{N}$, where $x_i \in \Re^n$ is the *i*-th input vector with known binary output label $Y_i \in \{-1, +1\}$. Then, the classification function is specified by:

$$Y_i = f(x_i) = w^T \phi(x_i) + b \tag{3}$$

where ϕ: $\Re^n \rightarrow \Re^m$ is the feature mapping of the input space to a high dimensional feature space. The data points become linearly separable by a hyperplane defined by the pair ($w \in \Re^m, b \in \Re$) (Vapnik, 1995). The optimal hyperplane separating the data is expressed as Eq. (5):

$$\begin{aligned} &\text{Minimize} \quad \Phi(w) = \|w\|^2 / 2 \\ &\text{Subject to} \quad Y_i \left[w^T \varphi(x_i) + b \right] \geq 1 \qquad i = 1, ..., N \end{aligned} \tag{4}$$

where $\|w\|$ is the norm for a normal weights vector of a hyperplane. This constrained optimization problem is solved by the following primal Lagrangian form:

$$L(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i \left[Y_i (w^T \phi(x_i) + b) - 1 \right] \tag{5}$$

where α_i are the Lagrange multipliers. Applying the Karush-Kuhn-Tucker conditions, solutions for the dual Lagrangian problem, α_i^0, decide the parameters w_0 and b_0 of the optimal hyperplane. Next, the decision function is generated by Eq. (7):

$$d(x_i) = \text{sgn}\left(w_0^T \phi(x_i) + b_0 \right) = \text{sgn}\left(\sum_{i=1}^{N} \alpha_i^0 Y_i K(x, x_i) + b_0 \right), \qquad i = 1, ..., N \tag{6}$$

The $K(x,x_i)$ is the kernel function and should satisfy Mercer's condition, as mentioned previously. In addition, the value of the kernel function is equal to the the inner product of two vectors x and x_i in the feature space $\phi(x)$ and $\varphi(x_i)$. Figure 3 shows the results of nonlinear SVM with the RBF kernel function.

Deep belief network–This study adopts DBN method to solve the seasonal time series data. A DBN is a feedforward neural network with many hidden layers. The initialize the weight (*w*) matrix and biases (*b*) usually adopts RBMs to training of a sequence of RBMs. The construct of RBMs is two layers which stochastic input dataset (Binary) are connected to stochastic output dataset. The figure 1 shows the first layer corresponds to input (Visible units *v*) and the second layer to the hidden units h of the RBMs (Chao et al., 2011).

Let v_i and h_j represent the states of visible unit *i* and hidden unit *j* respectively, and w_{ij}=w_{ji} is the bidirectional weights. The state probability of the units are

$$p_{vi} = 1 / \left[1 + e^{-\sum_j w_{ij} h_j} \right] \tag{7}$$

$$p_{hi} = 1 / \left[1 + e^{-\sum_j w_{ij} v_j} \right] \tag{8}$$

The procedure of RBMs is firstly training sample set to produce v_i. Then the hidden units are sampled according to probability of hidden unit. Repeating this process once more to update the states of visible unit and hidden unit produce one-step "reconstructed" state of visible unit and hidden unit. The update equation can be formulated as following:

$$\Delta w_{ij} = \varepsilon(< v_i h_j > - < v_i h_j >_{recon})$$

where ε is the learning rate. Similarly, the learning rules for the bias terms are respectively

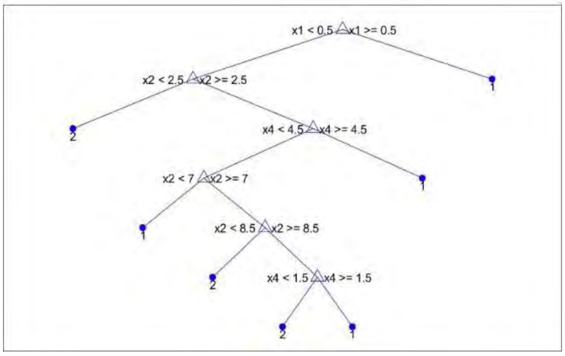
$$\Delta b_{ij} = \varepsilon(< v_i > - < v_i >_{recon})$$

The DBN can be constructed by stacking the RBMs one by one after training multiple RBMs individually. To construct a DBN we train sequentially as many RBMs as the number of hidden layers in the DBN. These RBMs are placed one on top of the other resulting in a DBN without the output layer (Chao et al., 2011).

Numerical example

In this study, e-learning datasets in 2019 was collected from the database. After clearing the data, 198 records were adopted. This research will develop methods that use actual e-learning platform data for testing verification. The six conditional attributes are "gender", "birth month", "public and private university", "home postal code", and decision variables “category". The gender and curriculum significant correlation, the main reason is that women's customer groups take business as the main course, and men's as the public subject. It can be seen from the recent data that the proportion of men who take online courses is 81.8%. The main target of online learning courses is Male, and engineering students are given priority. This research collaboration uses a decision tree method to clarify customer characteristics. Figure 1 presents a decision tree rule diagram and the five rules generated. It is found that it can provide marketing department reference information (1) male students almost selected engineering courses, and (2) female students almost selected business courses. Mainly, except those who live south of Changhua and were born after March or students who were born after September and are "non-Taipei". (3) Students from public or private universities will not affect the students' willingness to study e-learning courses. Furthermore, the two classifiers are used to divide the training set into 148 and the test set to 50. The results show that the support vector machine has better training errors and test errors. Therefore, it is recommended to recommend a marketing system to support vector machines in online learning courses. Classification may have more accurate performance. Through this industry-academia collaboration study, the recommendation marketing system can introduce support vector machine classification for intelligent recommendation, which may be better than deep confidence network.

Picture 1: The results of decision tree for e-learning dataset.



Rule 1: If it is "Male" then the online course is in the "Engineering" category

Rule 2: If it is "female" and was born in January and February, the online course is in the "business" category

Rule 3: If it is "female" and was born after March and lives "South of Changhua" then the online course is in the "Engineering" category

Rule 4: if it is "female" and was born in July and August and lives "north of Taichung" then the online course is in the "business" category

Rule 5: If it is "female" and was born after September and lives in "Taipei" then the online course is in the "business" category

Rule 5: If it is "female" and was born after September and lives in "non-Taipei area" then the online course is in the "engineering" category

Table 1: The results of classfication with SVM and DBN

Methodologies	Train error (No. of 148)	testing error(No. of 50)
SVM	0.2297	0.16
DBN	0.25	0.2

*Error is percentage of misclassification

Conclusion

Deep Belief Networks has been widely used in classification problems. This study designed a three-layer restricted Bozman machine network structure to train customer data. Energy functions and gradient training were used in the network. Customer data from the actual database were used. Experiment and provide digital marketing through decision tree, support vector machine, deep confidence network classifier for "gender", "birth month", "public and private university", "home postal code", and decision variable "class of study" The results of the systematic rules and classifier training found that male students mainly participated in online engineering courses, and female students were mainly business students, except for students who lived south of Changhua and were born after March or students who were born after September and "non-Taipei" , Students from public or private universities will not affect the students ’willingness to study courses digitally, and the proportion of study visits is similar. At the same time, it is found that classification by support vector machines may have a more accurate performance of deeper confidence networks, which is provided to manufacturers and platform developers.

Acknowledgements

This research was funded by the Ministry of Science and Technology of the Republic of China, Taiwan, grant numbers MOST-108-2221-E-029-020, MOST 107-2622-E-262-005–CC3, and MOST-107-2622-E-029-010-CC3.

REFERENCES

Chao, J., Shen, F., & Zhao, J. (2011). Forecasting exchange rate with deep belief networks. Proceedings of International Joint Conference on Neural Networks, San Jose, California, USA, 1259-1266.

Chen, T.-H. (2020). Do you know your customer? Bank risk assessment based on machine learning. *Applied Soft Computing*, 86, 105779.

Dursun, A., & Caber, M. (2016). Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis. *Tourism Management Perspectives*, 18, 153-160.

Quinlan, J.R. (1986). Induction of Decision Trees. *Machine Learning*, 1, 81-106.

Vapnik VN. (1995). *The Nature of Statistical Learning Theory*. New York: Springer-Verlag.

Vapnik, V., Golowich, S., & Smola, A. (1996). Support vector machine for function approximation, regression estimation, and signal processing. *Advances in Neural Information Processing Systems*, 9, 281–287.