Carry Distance of Moving Object Performance Comparison using Multiple Linear Regression and Artificial Neural Network

Jun Young Ko, Kyeong-Rok Kim, and Jae-Hyun Kim
Department of Electrical and Computer Engineering
Ajou University
Suwon, Korea
{kdb2658, nowhere1104, jkim}@ajou.ac.kr

***Abstract*— In this paper, carry distance of golf ball is calculated using data mining techniques that are Multiple Linear Regression (MLR) and Artificial Neural Network (ANN). The data mining models have six variables measured from observed data. One of the variables is dependent variable (carry distance of golf ball), the others are independent variables (club speed, attack angle, golf ball speed, launch angle, spin rate). Root Mean Square Error (RMSE) is used for performance evaluation between predictor value from data mining models and target value from observed data. The RMSE of MLR and ANN model is 2.12 and 0.77, respectively.**

Keywords— golf ball carry distance; Multiple Linear Regression (MLR); Artificial Neural Network (ANN); Root Mean Square Error (RMSE)

I. INTRODUCTION

Carry distance of golf ball is an important indicator because it decides golfer's shot accuracy. In [1], carry distance of moving object was calculated using the laws of physics such as drag force and lift force. However, it was not only complicated equation but also low accuracy compared to real value. When the relationships between variables is too complicated, a data mining technology explores hidden pattern and relationships between variables from big data [2]. The data mining technology is the process of discovering actionable information from big data. A data mining models are the classification and regression such as artificial neural network, genetics algorithm, k-nearest neighbor algorithm, support vector machine, decision trees [3].

In this paper, carry distance of golf ball is calculated using Multiple Linear Regression (MLR) and Artificial Neural Network (ANN) of data mining technologies. The MLR model structure has the predicted value (carry distance of golf ball) that is a linear combination of the parameters (regression coefficients) and explanatory variables (club speed, attack angle, golf ball speed, launch angle, spin rate). The ANN is a learning model that adjusts the parameters (weight and bias) each neuron to a target value (carry distance of golf ball) from a given input value (club speed, attack angle, golf ball speed, launch angle, spin rate). Root Mean Square Error (RMSE) is used in order to compare the performance of MLR and ANN.

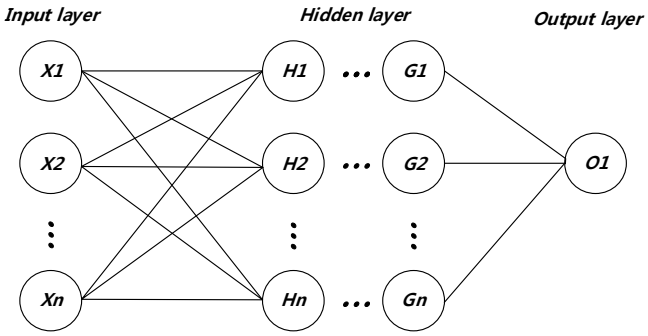

Fig. 1. The Artificial Neural Network model strructure

II. DATA MINING MODELS

A. Multiple Linear Regression (MLR)

MLR is one of the regression methods and it is a linear approach to model the relationship between a dependent variable and independent variables. The relationship is modeled using linear predictor functions that estimate unknown model parameters from the data. MLR model can be expressed as (1).

$$Y_i = \alpha + \beta_1 X_{i,1} + \cdots + \beta_n X_{i,n} \ , \qquad (1)$$

where α , β_n , $X_{i,n}$ denote the intercept, the regression coefficients of *n*-th explanatory variable, and *n*-th explanatory variable measured for the *i*-th observation.

B. Artificail Neural Network (ANN)

ANN is a biologically inspired computational model. It consists of artificial neurons, connected with coefficients (weights) which increase or decrease by strength of the signal of input value. Each neurons transmits a signal from one to another. An ANN model architecture has the input/output layer which has input value (independent variables) and the output value (dependent variable). It also has the hidden layer which has an activation function such as tanh, sigmoid, and ReLu function. In figure 1, we calculates the sum of the product of the input value and the weight using the ANN model. Then the value of the sum goes through the process of outputting by the activation function in the hidden layer. The output value of the hidden layer can be expressed as (2).

$$H_j = f_a\left(\sum_{i=1}^{n} X_i W_{ij} + \beta_i\right), \quad (2)$$

where f_a is activation function, X_i is input variable measured for i^{th} observation, W_{ij} is connection strength (weight) between the i^{th} input value and the j^{th} output value, β_i is bias. The ANN model uses the back propagation algorithm [4]. It changes the value of the weight and bias to minimize the error between the predicted and target value.

III. Results

The data were gathered from Trackman equipment which has accuracy of less than 1% error with real value. The shots of golfer were measured 4335 times indoors with 7iron. Table 1 shows the mean and standard variation of explanatory value (club speed, attack angle, golf ball speed, launch angle, spin rate) and target value (carry distance of golf ball). The MLR model is fitted using the Ordinary Least Squares (OLS). The OLS determines the regression coefficients of linear function by minimizing error sum of squares between dependent variable in observed data and predicted value of linear function. The equation of MLR model is derived as (3).

$$\begin{aligned} Y = &-70.6488 + 0.1004 * X_1 - 0.2861 * X_2 \\ &+1.8195 * X_3 + 1.0889 * X_4 - 0.0031 * X_5 , \end{aligned} \quad (3)$$

where Y is predicted value of linear function. X_1, X_2, X_3, X_4, and X_5 is club speed, attack angle, golf ball speed, launch angle, and spin rate from observed data. In order to evaluate performance of the model, RMSE between predicted value and target value was calculated. And the RMSE was 2.88.

The proposed ANN model structure has one input layer that has five nodes because it has five explanatory variables (club speed, attack angle, golf ball speed, launch angle, spin rate), one hidden layer that has an activation function each hidden nodes, and one output layer that has one predictor value. The sum of the product of the weight and the input value passed the hidden node in the hidden layer. To find out the best performance for a few hidden nodes, we change the number of hidden nodes from 2 to 15. The output value of the output layer is calculated as the sum of the activation function in the hidden nodes. Sigmoid function is used for activation function in this model that has output value between 0 and 1. Figure. 2 shows the RMSE between the target value and the predicted value according to the number of hidden nodes. The RMSE of hidden nodes 2 to 15 were measured as 2.18, 0.95, 0.93, 0.91, 0.91, 0.87, 0.84, 0.77, 0.93, 0.78, 0.89, 0.80, and 0.86, respectively. When the number of hidden nodes is 9, the RMSE is the lowest and the value is 0.77. On the other hands, when the number of hidden nodes is 2, RMSE is highest value that 2.18.

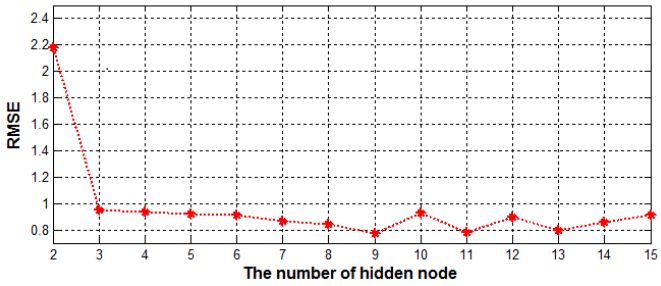

Fig. 2. RMSE between the target value and the predicted vlaue according to the number of hidden node

TABLE 1. The Mean and standard deviation of the variables

Variables	Mean	Std. deviation
Club speed (mph)	73.33	8.17
Attack angle (degree)	-2.19	1.94
Golf ball Speed (mph)	97.52	12.34
Launch Angle (degree)	19.04	3.64
Spin rate (rpm)	5911	1337
Carry distance (m)	117.22	22.40

IV. Conclusion

We calculated the golf ball of carry distance using data mining model that the MLR and the ANN. The models were constituted by five explanatory variables (club speed, attack angle, golf ball speed, launch angle, spin rate) and target value (carry distance of golf ball). The predicted value of data mining models are compared with target value using RMSE method. As the result, the ANN model has a minimum RMSE of 0.77 and better performance than the MLR model (RMSE = 2.88).

References

[1] J. Y. Ko, K. R. Kim and J. H. Kim, “Golf Ball Flight Trajectory Calculation Algorithm,” in Proc. KICS ICC 2018, High1 Resort, Korea, January 2018, pp. 637-638.

[2] B.M. Ramageri, “Data mining techniques and applications,” IndianJournal of Computer Science and Engineering, vol. 1, no. 4, December 2010, pp. 301-305.

[3] M. Kayri, I.kayri and M. T. Gencoglu, “The performance comparison of Multiple Linear Regression, Random forest and Artificial Neural Network by using photovoltaic and atmospheric data,” International Conference on Engineering of Modern Elctric Systems (EMES), Oradea, 2017, pp. 1-4.

[4] M. Kayri, “An Intelligent Approach to Educational Data: Performance Comparison of the Multilayer Perceptron and The Radial Basis Function Aritificial Neural Networks,” Educational Sciences: Theory & Practice, vol. 15, no.5, December 2015, 1247-1