A Proposed Artificial Neural Network for Predicting Movies Rates Category

Ibrahim M. Nasser, Mohammed O. Al-Shawwa, Samy S. Abu-Naser

Department of Information Technology, Faculty of Engineering and Information Technology
Al-Azhar University- Gaza, Palestine
Azhar.ibrahimn@gmail.com
mohammed.o.alshawwa@gmail.com

Abstract: We proposed an Artificial Neural Network (ANN) in this paper for predicting the rate category of movies. A dataset used obtained from UCI repository created for research purposes. Our ANN prediction model was developed and validated; validation results showed that the ANN model is able to 92.19% accurately predict the category of movies’ rate.

Keywords: Data Mining, Classification, Predictive Analysis, Artificial Neural Networks, movies classification

1. INTRODUCTION

Artificial neural networks (ANNs) are, similar to our neural networks and offer a relatively good technique, which solves the problem of classification and prediction. ANN is a collection of mathematical models, which can simulate characteristics of biological neural systems and have likenesses with adaptive human learning. ANNs made of connecting processing elements called neurons, connected by links, which contain weight coefficients that are, playing the role of synapses in our neural system. The neurons often come in three layers: input layer, one or more hidden layers and output layer, (ANN Architecture is shown in figure (1)). ANNs handle data as biological neural networks, in addition, ANN has the possibility of recalling, learning and eliminating errors, and high speed of getting the solution, [1] because of that, the neural networks can be used for solving complex problems, like classification and prediction [2]. ANNs were effectively applied in variety of applications for solving difficult and real problems [3].

ANN were found to be more efficient and more accurate than other classification techniques [4]. Classification by a neural network is done in two separate phases. First, the network is trained on a dataset. Then the weights of the connections between neurons are fixed so the network is validated to determine the classifications of a new dataset [5] . In this paper, we used about 70% of the total sample data for network training, and 30% for network validation.

While many models of ANNs have been proposed, the feed-forward neural networks (FNNs) are the most common and broadly used in many applications. Mathematically, the problem of training an FNN is the minimization of an error function E; In another word, to find a minimizer w = (w₁, w₂, . . . , wₙ) such that w = min E(w), where E is the batch error computed by the sum of square differences over all examples of the training dataset.

\[ E(w) = \sum_{p=1}^{P} \sum_{j=1}^{N_L} \left( j_p^L Y - t_{j,p} \right)^2 \]

\( j_p^L \) is the output of the j-th neuron that belongs to the L-th (output) layer, \( N_L \) is the number of neurons of the output layer, \( t_{j,p} \) is the anticipated response at the j-th neuron of the output layer at the input pattern p, and p represents the total number of patterns which used in the training dataset.

A traditional technique to solve this problem is by an iterative gradient-based training process, which produces a series of weights \( \{w_k\} \) starting from an initial point \( w_0 \in \mathbb{R}^n \) Using the iterative formula

\[ w_{k+1} = w_k + n_k d_k \]

where \( k \) is the current iteration, \( n_k > 0 \) is the learning rate and \( d_k \) is the decent search direction [5]. Our study main purpose is to develop a neural network as classification technique to predict the category of movies rate. A dataset from UCI repository [6,7] was used for this purpose.

Figure 1: ANN Architecture
1. Literature Review

There are many studies involving Artificial Neural Network (ANN) for example: Artificial Neural Networks and expert systems were employed to obtain knowledge for the learner model in the Linear Programming Intelligent Tutoring System (LP-ITS) to be able to determine the academic performance level of the learners in order to offer him/her the proper difficulty level of linear programming problems to solve[8-12,15,18,21-23]; for predicting the performance of a sophomore student enrolled in engineering majors in the Faculty of Engineering and Information Technology in Al-Azhar University of Gaza was developed and tested [37,45]; ANN model was developed and tested to predict temperature in the surrounding environment [20]; for predicting critical cloud computing security issues by using Artificial Neural Network (ANNs) algorithms. However, they proposed the Levenberg–Marquardt based Back Propagation (LMBP) Algorithms to predict the performance for cloud security level [32]; for predicting the MPG rate for the forthcoming automobiles in the foremost relatively accurate evaluation for the approximated number which foresight the actual number to help through later design and manufacturing of later automobile [17,36]; to predict efficiency of antibiotics in treating various bacteria types [40]; to predict the rate of treatment expenditure on an individual or family in a country [46], for detecting early-stage non-small cell lung cancer (NSCLC) [38]; for the diagnosis of hepatitis virus [34,41]; for predicting the Letters from twenty dissimilar fonts for each letter [35], for Email Classification Using Artificial Neural Network [14]; Classification Prediction of SBRC Ts Cancers Using Artificial Neural Network [16, 25]; for Diabetes Prediction Using Artificial Neural Network [29]; to predict Birth Weight [19]; to help cars dealers recognize the many characteristics of cars, including manufacturers, their location and classification of cars according to several categories including: Buying, Maint, Doors, Persons, Lug_boot, Safety, and Overall [13]; for Parkinson's Disease Prediction Using Artificial Neural Network [39,42,44]; for desktop PC Troubleshooting [27]; for Tomato Leaves Diseases Detection Using Deep Learning [26]; Plant Seedlings Classification Using Deep Learning [24,43]; for predicating software analysis and risk management [30,31].

2. Methodology

We got a movie ranking data set created by Mehreen Ahmed. We used this dataset which to build and validate our ANN model.

2.1 Dataset Description

Table 1: Original Dataset attributes description

<table>
<thead>
<tr>
<th>#</th>
<th>Attribute</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Movie</td>
<td>Text</td>
</tr>
<tr>
<td>2</td>
<td>Year</td>
<td>Integer</td>
</tr>
<tr>
<td>3</td>
<td>Genre</td>
<td>Integer</td>
</tr>
<tr>
<td>4</td>
<td>Gross</td>
<td>Integer</td>
</tr>
<tr>
<td>5</td>
<td>Budget</td>
<td>Integer</td>
</tr>
<tr>
<td>6</td>
<td>Screens</td>
<td>Integer</td>
</tr>
<tr>
<td>7</td>
<td>Sequel</td>
<td>Integer</td>
</tr>
<tr>
<td>8</td>
<td>Sentiment</td>
<td>Integer</td>
</tr>
<tr>
<td>9</td>
<td>Views</td>
<td>Integer</td>
</tr>
<tr>
<td>10</td>
<td>Likes</td>
<td>Integer</td>
</tr>
<tr>
<td>11</td>
<td>Dislikes</td>
<td>Integer</td>
</tr>
<tr>
<td>12</td>
<td>Comments</td>
<td>Integer</td>
</tr>
<tr>
<td>13</td>
<td>Aggregate Followers</td>
<td>Integer</td>
</tr>
<tr>
<td>14</td>
<td>Ratings</td>
<td>Real</td>
</tr>
</tbody>
</table>

2.2 Dataset Preprocessing and Transformation

We did some preprocessing and transformation so the data is fit for predictive analysis. We used the first 13 attributes as inputs to our model except movie, and year attributes. In addition, the Ratings attribute was the used as the class to be predicted based on the input attributes. We normalized the values of the attributes: gross, budget, screens, views, likes, dislikes, comments, and aggregate followers, so they became real because they were large integer numbers. Normalization formula was

\[
\text{new value} = \frac{(old value - \text{Min}(a_1 \ldots a_n))}{(\text{Max}(a_1 \ldots a_n) - \text{Min}(a_1 \ldots a_n))}
\]

We transform the class attribute (Ratings); we categorized rates based on the criteria showed below so it became nominal.

Table 2: Ratings Transformation

<table>
<thead>
<tr>
<th>Interval</th>
<th>Value</th>
<th>Transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – 2.9</td>
<td>Bad</td>
<td>1</td>
</tr>
<tr>
<td>3 – 4.9</td>
<td>Good</td>
<td>2</td>
</tr>
<tr>
<td>5 – 7.9</td>
<td>Very good</td>
<td>3</td>
</tr>
<tr>
<td>8 - 10</td>
<td>Excellent</td>
<td>4</td>
</tr>
</tbody>
</table>

The resulted dataset description is shown in table (3).
Table 3: Description after preprocessing

<table>
<thead>
<tr>
<th>#</th>
<th>Attribute</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Genre</td>
<td>Integer</td>
</tr>
<tr>
<td>2</td>
<td>Gross</td>
<td>Real</td>
</tr>
<tr>
<td>3</td>
<td>Budget</td>
<td>Real</td>
</tr>
<tr>
<td>4</td>
<td>Screens</td>
<td>Real</td>
</tr>
<tr>
<td>5</td>
<td>Sequel</td>
<td>Integer</td>
</tr>
<tr>
<td>6</td>
<td>Sentiment</td>
<td>Integer</td>
</tr>
<tr>
<td>7</td>
<td>Views</td>
<td>Real</td>
</tr>
<tr>
<td>8</td>
<td>Likes</td>
<td>Real</td>
</tr>
<tr>
<td>9</td>
<td>Dislikes</td>
<td>Real</td>
</tr>
<tr>
<td>10</td>
<td>Comments</td>
<td>Real</td>
</tr>
<tr>
<td>11</td>
<td>Aggregate Followers</td>
<td>Real</td>
</tr>
<tr>
<td>12</td>
<td>Ratings</td>
<td>Integer</td>
</tr>
</tbody>
</table>

2.3 The Neural Network

The resulted ANN Model is shown in figure (2).

2.4 Results

Our ANN model was able to predict the rate class with 92.19% accuracy, after 11776860 learning cycles with about 1% training error rate as seen in figure (3). In addition, Our Model showed that the most attribute that has effect on the movie rate was “sequel”. More details are shown in figure (4).

3. Conclusion

An artificial Neural Network for predicting the rate category of a movie was developed. The model was validated; it was 92.19 accurate in predict the rate category. This study showed that the neural network is able to predict movies rate category, so it can used for rating movies in the future.

References


5. K. D. P. P. Ioannis E. Livieris “Predicting students’ performance using artificial neural networks”.


