Prediction Model for Students' Future Development by Deep Learning and Tensorflow Artificial Intelligence Engine

Wilton W.T. Fok[1], Y.S. He[1], H.H. Au Yeung[1], K.Y. Law[1], KH Cheung[1], YY. Ai[1], P. Ho[1]
[1]The University of Hong Kong
e-mail: wtfok@eee.hku.hk

***Abstract*—Classification and prediction of students' performance in examination are the typical challenges for educators. Various traditional data mining methods such as decision tree and association rules were used to perform classification. In recent years, the rapid development of artificial intelligence and deep learning algorithm provided another approach for intelligent classification and result prediction. In this paper, a research on how to use Tensorflow artificial intelligence engine for classifying students' performance and forecasting their future universities degree program is studied. An appropriate and accurate forecast is important for providing prompt advice to student on program and university selection. For a more comprehensive consideration of an all rounded factors, the deep learning model analysed not only the traditional academic performance including Mathematic, Chinese, English, Physics, Chemistry, Biology and History, but also non-academic performance such as service, Conduct, Sport and Art. A few parameters in Tensorflow engine including the number of intermediate nodes and number of deep learning layers are adjusted and compared. With a data set of two thousands students, 75% of these data are used as the training data and 25% are used as the testing data, the accuracy ranged from 80% to 91%. The optimal configuration of the Tensorflow deep learning model that achieves highest prediction accuracy is determined. This study determined the factors affecting the accuracy of the prediction model.**

Keywords: e-Learning assessment, Artificial Intelligence, Deep Learning, prediction modelling

I. INTRODUCTION

How to predict students' performance is always a question concerned by the students' teachers and parents. Based on the past examination results and in-class assessments, it is possible to forecast the future development of the students. It is a challenging and important matters as it involves the large volume of data in educational databases and the result could impact the future development of a young kid. A good and accuracy prediction could bring the benefits and impacts to students, educators and academic institutions. Various type of data mining techniques had been used for performance prediction for decades, e.g. decision tree, Naive Bayes, K-Nearest Neighbor and Support Vector Machine [1]. However, with the rise of artificial intelligence and deep learning application, using AI engine such as Google Tensorflow for pattern recognition has now been rising it importance. In this paper, we will investigate how to use artificial intelligence and deep learning algorithm for pattern recognition and correlation of assessment results. There are some traditional data mining techniques that have been used to predict students' performance. Some researches educational data mining method had been done to identify those important attributes in students data.

II. TRADITIONAL METHOD FOR CLASSIFICATION AND PREDICTION

In order to build the predictive modeling, there are several traditional tasks used for example classification, regression and categorization. Algorithm such as association rules and decision tree are commonly used.

Association rule reflects the interdependence and correlation between different things. It is commonly used in physical stores and e-commerce recommendation systems. Likewise, it can be used for recommending schools and subjects to students. Support, confidence are two key concepts of them. In a given database, each transaction contains a set of items. Every rule is composed by two different itemsets X and Y. Support is the percentage of the transaction contains both X and Y, confidence is the percentage of Y that contains X. In other words, support is the probability while confidence is the conditional probability. If the minimum thresholds of them, which have to be set by us, are satisfied, there is an association that exists between X and Y. Association rule mining is more applicable to the situation where the index in the record are discrete value. If the indicator values in the original database are continuous data, and appropriate discretization should be performed previously.

The classification decision tree model is a tree structure that describes the classification of instances. The decision tree consists of nodes and directed edges. There are two types of nodes: internal nodes and leaf nodes, internal nodes represent a feature or attribute, and leaf nodes represent a class.

When categorizing, a certain feature of the instance is tested starting from the root node, and the instance is assigned to its child node according to the test result; at this time, each child node corresponds to a value of the feature. This recursively moves down until it reaches the leaf node, and finally assigns the instance to the class of the leaf node. The decision tree can be seen as a set of if-then rules: a rule is constructed from the root node of the decision tree to each path of the leaf nodes; the characteristics of the internal nodes on the path correspond to the conditions of the rules, and the leaf nodes Corresponds to the classification of the conclusions.

978-1-5386-6147-5/18/$31.00 ©2018 IEEE

III. New Methods Using Deep Learning and Artificial Intelligence

Neural network is another emerging technique used in educational data mining. The advantage of neural network is that it has the ability to detect all possible interactions between predictors variables [2]. When more computing power is nowadays available, more layers of neural network can be implemented and deep learning analysis can be practically implemented. Deep learning could perform detection even in a complex nonlinear relationship between dependent and independent variables [3]. It is considered as one of the best prediction method.

In this study, the Google Tensorflow Deep Learning analystic engine is used to predict students future development. The attributes analyzed are the academic performance such as the examination scores of various traditional academic subjects including Mathematic, Chinese, English, Physics, Chemistry, Biology and History, and their non-academic performance such as conduct, sport, arts and participation of services are included in the analysis. Two thousand records are generated according to the following rules:

1. *Good at physics and mathematics but poor in Chinese → engineering in University A*
2. *Good at physics, Chemistry and English but poor in Sport → Medicine in University A*
3. *Good at physics, Chemistry, English and Conduct but poor in Chinese → Medicine in University B*
4. *Good at Chinese and Service but poor in biology, maths and conduct → Education in University C*
5. *Good at Chinese but poor in biology, maths and conduct → Chinese in University C*
6. *Good at Chinese, English but poor in Maths, Physics → Translation in University D*
7. *Good at Sport and Chinese but poor in history → Education in University D*
8. *Good at Sport and Chinese and poor in conduct → Sport Science in University D*
9. *Good at Maths and art but poor in Chinese, English → design in University E*
10. *Good at Physics, Maths, English but poor in Chinese → engineering in University B*

At the beginning, these 2000 data sets were divided into two parts. The first data part contains 75% of the data set, i.e. 1500 data sets, were set as the training data, while the remained 25% data, i.e. 500 data sets, were used as the test data.

The assessment performance of various subjects were used as the input x_1, x_2,… x_n . The corresponding program and university selection outcome were used as the output y_1, y_2,… y_m . To build up the prediction model, the Convolution Neural Network (CNN) model is used. CNN is commonly used in the pattern recognition problem. Using a one-hot output system, a full connected network was built up with the relationship between different nodes is Wx + b = y. Applying the softmax cross entropy to evaluate its cost function, the main structural was completed as shown in Figure 1.

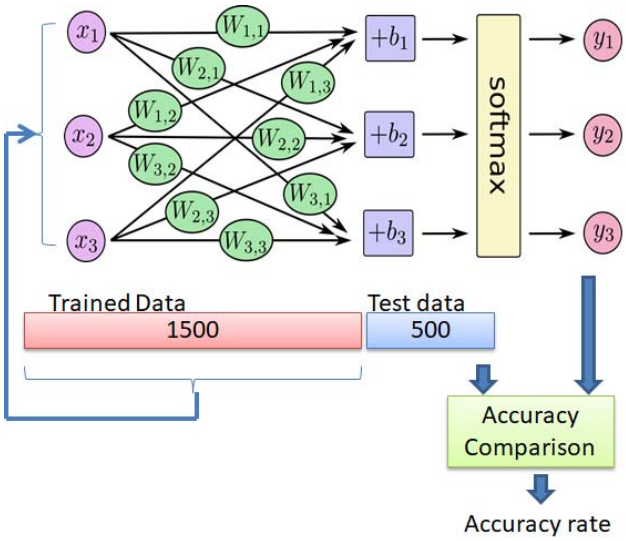

Figure 1. Structure of the Convolution Neural Network and softmax cross entropy

In this research, Python and the Deep Learning Engine Tensorflow is used for the development. TensorFlow is an open-source software library for dataflow programming across a range of tasks. It is a symbolic math library, and also used for machine learning applications such as neural networks. It enables this research to be started easily with deep learning in the cloud. The framework has broad support in the industry and has become a popular choice for deep learning research and application development, particularly in areas such as computer vision, natural language understanding and speech translation. The scores and future development of students follow certain pattern and therefore this pattern recognition engine is used for the analysis. The following loop for iteration is run to train the model:

```
for itt in range(epoch):
    avg_cost = 0
    feed_dict = {X: input, Y: output}
    c, _ = sess.run([cost, optimizer],
feed_dict=feed_dict)
```

After the model is trained, the following code in Tensorflow is run to test the performance of the trained model:

```
# Test model and check accuracy
correct_prediction =
tf.equal(tf.argmax(hypothesis, 1),
tf.argmax(Y, 1))

accuracy =
tf.reduce_mean(tf.cast(correct_prediction,
tf.float32))

result = sess.run(accuracy, feed_dict={X:
input_t, Y: output_test})

print('Accuracy:',result )
record = record + str(result) + ','
```

The number of hidden layers and number of nodes are adjusted for comparing the best performance configuration (Figure 2.). By adapting different calculation factor and element, such as hidden nodes, hidden layer and learning rate and so on, accuracy up to 90% was achieved.

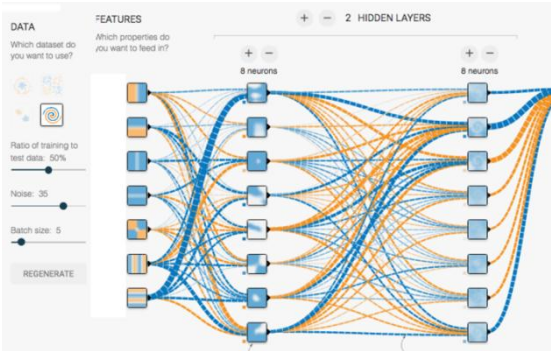

Figure 2. Hidden layers in Tensorflow

IV. RESULTS

A. *Effect of Number of Hidden Node and Layers to the Prediction Accuracy*

The number of hiddem nodes (j) and number of hidden layers of deep learning (k) are adjusted and the accuracy of the prediction is compared. The prediction accuracy against different j and k are summarised as follow:-

(j) (k)	10	20	30	40	50	60	70	80
1	0.84	0.86	0.91	0.89	0.88	0.88	0.88	0.84
2	0.80	0.85	0.85	0.83	0.84	0.83	0.85	0.86
3	0.67	0.80	0.84	0.83	0.84	0.85	0.83	0.83
4	0.15	0.83	0.83	0.80	0.86	0.85	0.83	0.85
5	0.15	0.87	0.77	0.82	0.84	0.80	0.85	0.83

The number of hidden node (j) and the number of hidden layers (k) and the prediction accuracy are plotted in the 3D chart in Figure 3. In this result, the number of hidden layers does not have significant impacts on the accuracy. However, when the number of hidden layers increase, the accuracy also increase until the number of hidden layer exceed 20.

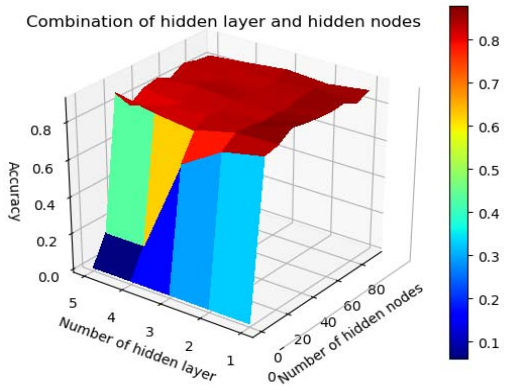

Figure 3. No. of hidden node (j) and No. hidden layers (k) Vs the prediction accuracy

B. *Effect of Number of Learning Rate and Number of Iteration to the Prediction Accuracy*

Next, the relationship between the number of iteration (p) and the learning rate (q) are compared. The number of hiddle layer (k) is set to 3 while the number of hidden nodes (j) is set to 20. The number of iteration (p) and the learning rate (q) are summarized in the table below and in Fig. 4.

(q) \ (p)	5000	10000	15000	20000	25000
0.001	0.834	0.842	0.836	0.828	0.824
0.002	0.83	0.828	0.836	0.836	0.802
0.003	0.78	0.834	0.844	0.818	0.856
0.004	0.862	0.846	0.838	0.86	0.844
0.005	0.834	0.838	0.788	0.856	0.846
0.006	0.818	0.858	0.842	0.816	0.848
0.007	0.492	0.858	0.852	0.818	0.856
0.008	0.862	0.844	0.854	0.828	0.856
0.009	0.806	0.836	0.838	0.696	0.832
0.010	0.808	0.858	0.842	0.828	0.822
0.011	0.802	0.856	0.86	0.83	0.866
0.012	0.846	0.826	0.834	0.646	0.852
0.013	0.84	0.82	0.854	0.854	0.862
0.014	0.862	0.846	0.832	0.652	0.83
0.015	0.824	0.826	0.878	0.528	0.85
0.016	0.76	0.822	0.84	0.838	0.848
0.017	0.844	0.832	0.154	0.83	0.858
0.018	0.838	0.852	0.808	0.844	0.82
0.019	0.864	0.828	0.236	0.838	0.782

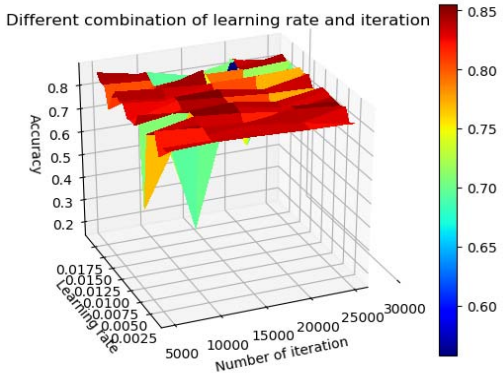

Figure 4. No. of iteration (p) and the learning rate (q) vs the prediction accuracy

V. CONCLUSIONS

Classification and prediction is a general problem. Traditional data mining techniques such as association rules, decision tree, clustering and so on had been use for a few decades for solving this problem. The raising popularity of using Tensorflow for deep learning and artificial intelligence opened a new approach and direction for solving classification problem and prediction of non-linear results.

In this research, the number of hidden layers, hidden notes, the number of iteration and the learning rate are adjusted and compared. It is discovered that it is not always true that the deep the deep learning model, i.e. the more number of hidden layer, the more accurate will the result be. There is an optimal point at are required to be tested and identified.

For the learning rate, a higher learning rate could help to speed up the convergence of the trained model. However, if the learning rate is too high, the result might overshoot the optimal point. Therefore, the prediction performance could be improved with a low momentum and high learning rate. Then gradually, the momentum can be increased and the learning rate can be decreased for ensuring convergence.

This study demonstrated that deep learning could be an effective tool for predicting the students' performance. The result ranged from 80% to 91%. The prediction result is good enough to provide appropriate recommendations for students, their teachers and parents to decide their development pathway. It is believed that more applications of deep learning could be used for education and corporate staff training in the future.

REFERENCE

[1] Amirah Mohamed, Shahiri Wahidah, "A Review on Predicting Student's Performance Using Data Mining Techniques", Procedia Computer Science, Volume 72, 2015, Pages 414-422, https://www.sciencedirect.com/science/article/pii/S1877050915036182

[2] G. Gray, C. McGuinness, P. Owende, An application of classification models to predict learner progression in tertiary education, in: Advance Computing Conference (IACC), 2014 IEEE International, IEEE, 2014, pp. 549–554.

[3] P. M. Arsad, N. Buniyamin, J.-l. A. Manan, A neural network students' performance prediction model (nnsppm), in: Smart Instrumentation, Measurement and Applications (ICSIMA), 2013 IEEE International Conference on, IEEE, 2013, pp. 1–5.

[4] Joonhyuck Lee , Dongsik Jang and Sangsung Park, "Deep Learning-Based Corporate Performance Prediction Model Considering Technical Capability" MDP http://www.mdpi.com/2071-1050/9/6/899/pdf

[5] Rianne Conijn, Chris Snijders, Ad Kleingeld, Uwe Matzat, "Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS. October 2016IEEE Transactions on Learning Technologies PP(99):1-1