Predicting Student Performance with Deep Neural Networks

Problem Statement

In present educational systems, student performance prediction is getting worsen day by day. Predicting student performance in advance can help students and their teacher to keep track of progress of a student. Many institutes have adopted continuous evaluation system today. Such systems are beneficial to the students in improving performance of a student. The purpose of continuous evaluation system is to help regular students.

In recent years, Neural Networks have seen widespread and successful implementations in a wide range of data mining applications, often surpassing other classifiers. This study aims to investigate if Neural Networks are a fitting classifier to predict student performance from Learning Management System data in the context of Educational Data Mining. To assess the applicability of Neural Networks, we compare their predictive performance against six other classifiers on this dataset. These classifiers are Naive Bayes, k-Nearest Neighbors, Decision Tree, Random Forest, Support Vector Machine and Logistic Regression and will be trained on data obtained during each course. The features used for training originate from LMS data obtained during the length of each course, and range from usage data like time spent on each course page, to grades obtained for course assignments and quizzes. After training, the Neural Network outperforms all six classifiers in terms of accuracy and is on par with the best classifiers in terms of recall. We can conclude that Neural Networks outperform the six other algorithms tested on this dataset and could be successfully used to predict student performance.

Background

In recent years, the use of internet-based educational tools has grown rapidly (Jordan, 2014) as well as the research surrounding them (see Figure 1). These tools provide a clear advantage for students and teachers alike, with the ability to access and share course data from anywhere in the world, track student progress and provide rich educational content.

Amrieh, et al. proposed a prediction model for students' performance based on data mining methods with some few features called student's behavioral features. The model was evaluated in three different classifiers; Naïve Bayesian, Artificial Neural Network and Decision tree. Random Forest, Bagging and Boosting were used as ensemble methods to improve the classifier's performance. The model achieved up to 22.1% more in accuracy compared when behavioral features were removed. It increased up to 25.8% accuracy after using the ensemble methods.

The family of classifiers this study focuses on, Neural networks, have shown promising results in domains like speech recognition (Graves &Jaitly, 2014), computer vision (Venugopalan et al., 2014), recognizing music (Costa, Oliveira, & Silla, 2017), playing complex games like GO (Wang et al., 2016) and economic forecasting (Nametala, Pimenta, Pereira, &Carrano, 2016), but their use in EDM has thus far been limited compared to other classification algorithms (Baker &Inventado, 2014). This can be partly explained by their difficulty to set up, the lack of convenient all in one packages that are easy to use and the often-long training times (Gaur, 2012). But they do offer clear benefits over other machine learning algorithms. They can classify instances in domains that are not linearly-separable and can handle noisy and complex data (Schmidhuber, 2015). These properties make them especially suited for a domain like EDM where the data, given the fact that it is based on human behavior, can be complex, might contain irrelevant entries as well as nonlinear relations.

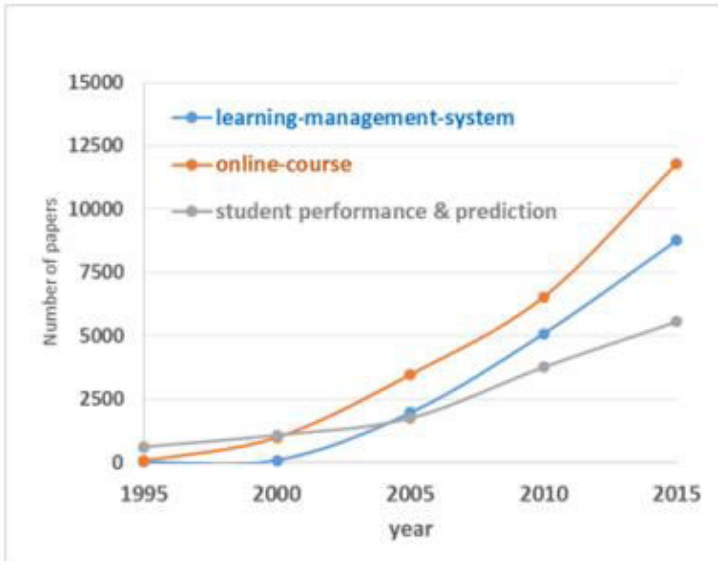

Figure 1: Number of papers in Educational Data Mining related fields. source: Google Scholar.

Methodology

Neural Networks in Student Performance Prediction

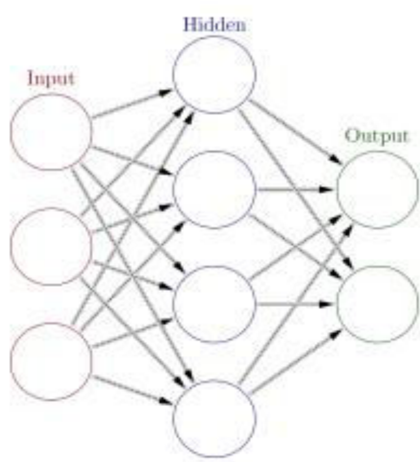

Figure 2: The structure of a Neural Network (Glosser.ca, 2013).

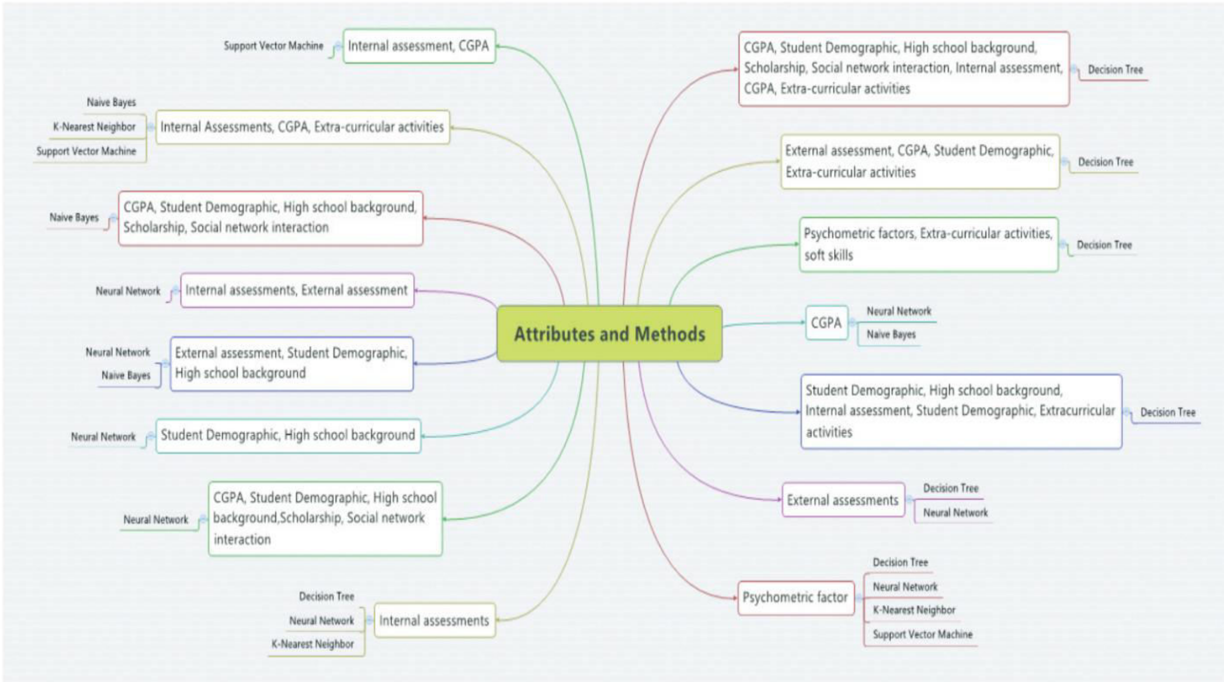

Fig 3: List of common attributes and methods used in predicting student's performance [2]

This study focuses on to predict student performance belongs to the family of Neural Networks. Neural Networks are algorithms that mimic the way our brain works. They consist of an array of interconnected nodes that exchange information among each other (see Figure 2), comparable to the way our neurons, connected by dendrites and axons, exchange information. They learn iteratively over time by observing different examples, similarly to how children can learn skills from their parents by observation. However, unlike children that 4 can learn recognize and object after only observing it once, Neural Networks often require a greater set of observations to attain sufficient predictive capacity. In this section the steps that were performed to extract the predictors are explained and the training process and parameter tuning for the classifiers are clarified: -

Dataset:

Link: https://www.kaggle.com/aljarah/xAPI-Edu-Data

Name: Students' Academic Performance Dataset

Dataset attributes: -

Students Dataset		
Name	**Data Type**	**Distinct Values**
Gender	Nominal	2
Nationality	Nominal	14
Place of Birth	Nominal	14
Stages	Nominal	3
Grades	Nominal	12
SectionID	Nominal	3
Topic	Nominal	12
ParentResponsible	Nominal	2
Semester	Nominal	2
Raised hand	Numeric	0-100
Visited Resource	Numeric	0-100
Viewing Announcement	Numeric	0-100
Discussion Group	Numeric	0-100
Parent Answering	Nominal	2
Parent Satisfaction	Nominal	2
Student Absent day	Nominal	2

Experimental Design

In this problem, we have to build a Deep Neural Network linear classifier model to predict the performance of students. This process should be followed once the dataset is preprocessed: data cleaning and data transformation. The DNN model will be built using python3 and tensorflow 1.3.0.

Artificial neural networks (ANNs) are parallel computational models comprised of densely interconnected, adaptive processing units, characterized by an inherent propensity for learning from experience and discovering new knowledge. Due to their excellent capability of self-learning and self-adapting, they have been extensively studied and have been successfully utilized to tackle difficult real-world problems (Bishop 1995; Haykin 1999) and are often found to be more efficient and more accurate than other classification techniques (Lerner et al. 1999). Classification with a neural network takes place in two distinct phases. First, the network is trained on a set of paired data to determine the inputoutput mapping. The weights of the connections between neurons are then fixed and the network is used to determine the classifications of a new set of data. Although many different models of ANNs have been proposed, the feedforward neural networks (FNNs) are the most common and widely used in a variety of applications.

Python is a full featured for general purpose programming language. It is a mature and fast expanding platform for scientific research and numerical computing. Python host numerous open source libraries and almost all general-purpose libraries for machine learning which can be further use for deep learning models. All this benefits from the python ecosystem lead to the top two libraries for numerical analysis of deep learning was developed for python language, that is Tensorflow and Theano library.

TensorFlow is an open source library for computing numerical using data flow graphs. The data flow graphs are also known as Static Computation graph. A developer must first design the input layer and connect every input layer to the hidden layer then the same from hidden layer to output layer. The graphs are made of tensors and ops, defining all the neural networks and all mathematical calculations. The session helps to run the graph. Tensorflow comes with Graphical Processing Unit package where all the matrix calculations can be done efficiently and much faster. Once data is preprocessed, the data is divided into two parts training and testing dataset. It is divided in the ratio 3:1 (Train/Test). In training dataset, the features and classes are split and stored in a tensorflow placeholder. Both training dataset classes records are One-hot encoded, it is a process where class variables are converted into a numerical form that will be provided to deep neural network model for effective prediction.

Other approaches: -

- Network-Based Clustering
- Baseline Methods

References

1. http://delivery.acm.org/10.1145/640000/638180/s2.pdf?ip=115.249.53.74&id=638180&acc=OPEN&key=045416EF4DDA69D9%2E0B5B0253775BA20F%2E4D4702B0C3E38B35%2E6D218144511F3437&__acm__=1525156531_876a762e556fd775eba267a77d75e7e0
2. https://ac.els-cdn.com/S1877050915036182/1-s2.0-S1877050915036182 main.pdf?_tid=44012f66-b408-473b-a41b-69790c3b51e2&acdnat=1525166323_ac1e1a0715d091a4ed3f38939666bfd4