Perspectives in Science (2016) 8, 364—366

ScienceDirect
journal homepage: www.elsevier.com/pisc

Towards the integration of multiple classifier pertaining to the Student's performance prediction☆

Mrinal Pandey*, **S. Taruna**

Banasthali University, Jaipur, Rajasthan, India

Received 6 February 2016; accepted 12 April 2016
Available online 27 April 2016

KEYWORDS
Decision tree;
K-nearest neighbour;
Aggregating one-dependence estimators (AODE) predictions;
Algorithms

Summary Accurate predictions of students' academic performance at early stages of the degree programme helps in identification of the weak students and enable management to take the corrective actions to prevent them from failure. Existing single classifier based predictive modelling is not easily scalable from one context to another context, Moreover, a predictive model developed for a particular course at a particular institution may not be valid for a different course at the same institution or any other institution. With this necessity, the notion of the integrated multiple classifiers for the predictions of students' academic performance is proposed in this article. The integrated classifier consists of three complementary algorithms, namely Decision Tree, K-Nearest Neighbour, and Aggregating One-Dependence Estimators (AODE). A product of probability combining rule is employed to integrate the multiple classifiers for the prediction of academic performance of the engineering students. This approach provides a generalized solution for student performance prediction. The proposed method has been applied and compared on three student performance datasets using t-test. The proposed method is also compared with KSTAR, OneR, ZeroR, Naive Bayes, and NB tree classifiers as well as with the individual classifiers.

Introduction

In the modern era of data mining, researchers are continually advocating for the use of multiple classifiers to solve classification problems. The concept of combining classifier is based on the assumption that the different classifiers, which uses a different data representation, different concept and different modelling techniques are most likely to arrive at classification results with different patterns of

☆ This article belongs to the special issue on Engineering and Material Sciences.

* Corresponding author.
E-mail address: mrinalpandey14@gmail.com (M. Pandey).

http://dx.doi.org/10.1016/j.pisc.2016.04.076

generalization (Kotsiantis and Pintelas, 2005). Researchers have proved that integration of diverse classifiers reduces the classification errors (Kotsiantis and Pintelas, 2005). The objective of this research is to devise a flexible, scalable and generalized predictive modelling approach for students' performance prediction.

Related work

A number of related works based on mathematical and prediction model have been developed for students' performance prediction. Hien and Haddawy (2007) defined a case based mechanism from the Bayesian network to perform the predictions. In another study (Kotsiantis, 2012) a DSS based on regression techniques was proposed for the forecasting the student's grades. A simulation tool (Saxena, 2012) based on Fuzzy System, Neural Network and Genetic Algorithm was developed for the classification and analysis of the students' performance. In Livieris et al. (2012), a neural network was used for the development of a student performance prediction tool particularly for mathematics course that enables educators to identify weak students, was also developed. A decision support system for predicting student performance was proposed using the Naïve Bayes algorithm in Lalit Dole and Jayant Rajurkar (2014). An ensemble based algorithm has been proposed for the students' performance prediction in near the beginning stages to minimize the failure rate by counselling the risk associating students (Sharaf et al., 2013). A comparative study of homogenous ensemble has been performed in Pandey and Taruna (2014a) for students' academic performance prediction. A review of data mining techniques related to students' performance prediction is presented in Shahiri et al. (2015).

Methodology

The methodology adopted for this research starts with the data collection followed by initial pre-processing, attribute selection and balancing the class. The next phase is followed as a model construction. In this step three complementing classifier namely DT (J48), KNN (IBK) and AODE are integrated and proposed a single composite model (KNNAD) based on voting strategy.

Dataset

The data set used for this study was from an engineering college in India. The datasets consist of academic information as well as the demographic information of the undergraduate engineering students. Three student performance datasets were used for the study. The dataset consist of complete (1000), complete (525) and outliers (960) instances, respectively. However, the dataset outliers (960) are a filtered dataset or subset of the complete (1000) dataset, which was obtained from the research (Pandey and Taruna, 2014b) after removing the outliers. All three datasets contained the same number of attributes.

Pre-processing

The initial pre-processing was conducted using WEKA (Hall et al., 2009). Thereafter chi-square based on a ranker method was used for attribute selection. The eight higher ranker attributes amongst 18 attributes were considered for all three datasets. A combination of two class-balancing techniques, under sampling followed by SMOTE (Chawla et al., 2002) is used to rebalance the dataset.

Model construction

In order to construct the integrated model, three algorithms, namely K-Nearest Neighbour (IBK), AODE (Webb et al., 2005) and Decision Tree (J48) (Quinlan, 1993) were used as base classifiers. In present research the DT classifier is used for better visual representation, while KNN classifier is selected due to its better performance for large size data sets. On the other hand AODE (Advance Version of Naive Bayes) overcomes the limitation of overfitting problem of DT, as it is a linear classifier and less likely to suffer from an overfitting in case of large data sets also. The AODE and KNN algorithms complement the DT algorithm. Additionally, these three aforementioned algorithms are used to enhance the globalization and flexibility for different types of data sets. All three aforementioned classifiers were integrated using the product of probability voting rule.

$$f_j(x) = \prod_{n=1}^{N} V_{n,j}(x)$$

Where n: Number of classifiers and j: Class.

In this experiment, each individual classifier (IBK, AODE and J48) generates their hypothesis h1, h2 and h3, respectively. For each output class, a posteriori probability is generated by the individual classifier, which is multiplied to find the product of probabilities and the class is represented by the maximum of a posterior probability. This posterior probability is selected to be the voting hypothesis (h*) for the final decision. Fig. 1 shows the proposed model.

Experiments & results

Initially proposed model is compared according to t-test with its individual classifiers such as AODE, KNN (IBK) and J48 decision tree as well as with other five different representative classifiers namely Naïve Bayes, KSTAR, OneR, ZeroR and Naïve Bayes Tree. The results of Table 1 is remarkable, the accuracy of the proposed model (KNAND) is better than five classifiers namely AODE, NB, ZeroR, OneR and NBT for all three datasets, while there is no significant difference in the accuracy of the KSTAR lazy learner and proposed voting

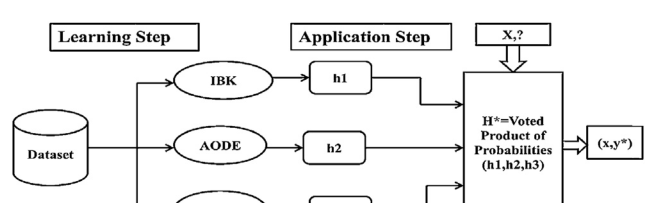

Figure 1 Proposed model: KNNAD.

Table 1 Performance accuracy comparisons of proposed model.

Data set	Proposed voting	AODE	IBK	J48	NB	KSTAR	ZeroR	OneR	NBT
Complete (1000)	87.03	81.85 *	87.28	85.25 *	80.63 *	85.66	38.25 *	82.41 *	85.01 *
Complete (525)	86.76	80.93 *	87.23	84.55	72.59 *	85.66	34.97 *	62.05 *	83.77 *
Filtered (960)	98.86	93.03 *	98.96	98.86	91.57 *	98.96	59.31 *	84.19 *	96.98 *

Table 2 Experimental statistics.

Data set	Correctly classified instances (%)	Incorrectly classified instances (%)	Kappa statistic	Mean absolute error	Root mean squared error	Relative absolute error (%)	Time taken in (seconds)
Complete (1000)	87.84	12.16	0.83	0.07	0.23	17.9	0.02
Complete (525)	90.35	9.65	0.87	0.05	0.22	13.8	0.02
Filtered (960)	99.01	0.99	0.99	0.01	0.07	1.4	0.02

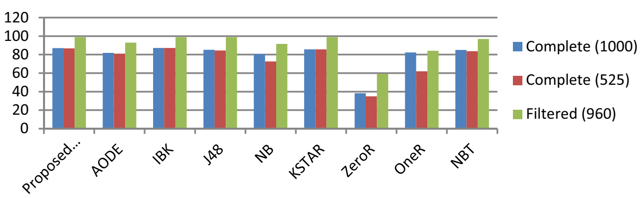

Figure 2 Performance accuracy comparisons for three data sets.

model, however, the J48 decision tree algorithm is less accurate for one dataset consisting of 1000 instances but there is no significant difference for other two data sets.

Table 2 represents the detailed experimental statistics for the proposed model for all the three datasets, as correctly and incorrectly classified instances, kappa statistic, mean absolute error, root mean squared error, relative absolute error and time taken to build an ensemble model.

The Fig. 2 depicts the accuracy of the classifiers for three data sets. It is clear from the figure that the accuracy is highest for filtered dataset. It can also be observed that the performance of the proposed algorithm is highest, while the performance of the zero is worst among all eight classifiers for all three datasets.

Conclusion & future work

In this research a heterogeneous multiple classifier-based framework is presented, which integrates three classifiers AODE, IBK and J48 using the voting methodology and proposed a single composite model. This integrated model was constructed and tested against three sets of students' datasets and results reflected as the consistent behaviour and performance accuracy.

However, the model has been proposed for predicting the academic performance of the students, particularly in engineering discipline, but this can also be applicable for other domains of data mining applications as a future work. It can also be used for the development of decision support system.

References

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 32, 1—357.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, Ian H., 2009. The WEKA data mining software: an update. SIGKDD Explor. Newsl. 11 (1).

Hien, N.T.N., Haddawy, P., 2007. A decision support system for evaluating international student applications. In: In Proceedings of the 37th ASEE/IEEE Frontiers in Education Conference.

Kotsiantis, S.B., Pintelas, P., 2005. Local voting of weak classifiers. Int. J. Knowl-Based Intell. Eng. Syst. 9, 239—248, 239, IOS Press (2005).

Kotsiantis, S.B., 2012. Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades. Artif. Intell. Rev. 37 (4), 331—344.

Lalit Dole, Jayant Rajurkar, December 2014. A decision support system for predicting student performance. Int. J. Innovat. Res. Comput. Commun. Eng. 2 (12).

Livieris, I.E., Drakopoulou, K., Pintelas, P., 2012. Predicting students' performance using artificial neural networks. In: In Proceedings of the 8th Pan-Hellenic Conference Information and Communication Technology in Education.

Pandey, M., Taruna, S., October 2014. A comparative study of ensemble methods for students' performance modelling. Int. J. Comput. Appl. 103 (8), 26—32.

Pandey, M., Taruna, S., 2014b. A multi-level classification model pertaining to the students' academic performance prediction. Int. J. Adv. Eng. Technol. 7 (4), 1329.

Quinlan, J.R., 1993. C4. 5: Programs for machine learning. Morgan Kaufmann, San Francisco, CA, USA.

Saxena, U.R., 2012. Integrating neuro-fuzzy systems to develop intelligent planning systems for predicting students' performance. Int. J. Eval. Res. Educ. 1 (2), 61—66.

Shahiri, A.M., Husain, W., Rashid, N.A., 2015. A review on predicting student's performance using data mining techniques. In: In Proceedings of the Third Information Systems International Conference 2015, 72, pp. 414—422, http://dx.doi.org/10.1016/j.procs.2015.12.157.

Sharaf, A., Malaka, E.D., Moustafa, A., Harb, H.M., Abdel, H.E., 2013. Adaboost ensemble with simple genetic algorithm for student prediction model. Int. J. Comput. Sci. Inf. Technol. 5.

Webb, G.I., Boughton, J.R., Wang, Z., 2005. Not so naive Bayes: aggregating one-dependence estimators. Mach. Learn. 58 (1), 5—24.