Rodrigo, M. M. T. et al. (Eds.) (2021). Proceedings of the 29th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Explore the Contribution of Learning Style for Predicting Learning Achievement and Its Relationship with Reading Learning Behaviors

Fuzheng ZHAO[a], Bo JIANG[b], Juan ZHOU[c] & Chengjiu YIN[d*]
[a]*Graduate School of System Informatics, Kobe University, Japan*
[b]*Department of Educational Information Technology, East China Normal University, China*
[c]*School of Environment and Society, Tokyo Institute of Technology, Japan*
[d]*Information Science and Technology Center, Kobe University, Japan*
*yin@lion.kobe-u.ac.jp

Abstract: Prediction is an important branch of research in learning analytics, in which the prediction of learning achievement has much practical value for improving instructional management and enhancing learning effectiveness. As a type of cognitive data, students' learning style data offers great potential for predicting their learning achievement. Based on the analysis of the contribution of learning style data on prediction model creation, this study uses the Felder and Silverman learning style scale to examine 238 students' learning styles as feature elements and explores the feature importance using six machine learning algorithms to create models for learning achievement prediction. Besides, to identify the relationship between learning styles and learning behaviors, and the hidden learning patterns behind learning styles, the study collected reading log data using the E-book system for correlation and principal component analysis. It was found that the Decision Tree model obtained the best results in terms of accuracy and other indicators. Secondly, the VisualScore feature showed the greatest influence on all the six models used. Thirdly, the study also found that learning styles were highly correlated with repeated learning and marking behavior in reading behavior. Finally, the analysis showed that the visual and verbal dimensions under the VisualScore features had three common learning patterns of repeated reading, marking, and mobile reading, in addition to differences in learning patterns in terms of time spent.

Keywords: Learning style, learning prediction, reading learning behavior

1. Introduction

In the e-publication era, learning analytics provides a huge analysis and mining potential (Zhao, Huang, & Yin, 2018) to rethink the role (Yin et al., 2019) and strategy of education technology in the learning practice (Yin & Huang, 2018). Various predicted methods were selected to detect whether they have the ease and effectiveness of predicting effects (Huang et al., 2020). Learning style is mainly expressed by preference in learning methods. It is worth noting that learning style has a certain correlation with cognitive ability, but it has no absolute relationship with the strength or weakness of ability (Hames & Baker, 2015).

This study experimented to collect the data by the Index of Learning Styles and reading learning behavior by E-book system, to understand the contribution of learning style data to predicting achievement, figure out the most optimal features, as well as explore the relationship between learning style data and reading learning behavior.

2. Methodology

The learning style data including four variables, such as ActiveScore, SensingScore, VisualScore, and SequentialScore, was collected from 238 participants by the Index of Learning Styles questionnaire. In this study, the Scikit-learn was used to make model creation, such as Decision Tree (DT), Random

Forests (RF), XGBoost (XG), Logistic Regression (LR), Support Vector Machines (SVM), and K-nearest Neighbors (KNN). Subsequently, as a binary classification case, Accuracy, F-score, Recall, Precision, and AUC meet the requirements of evaluation for model performance. Finally, we used impurity-based feature importance, coefficients feature importance, and permutation feature importance to calculate feature importance.

To analyze the relationship between reading learning behavior and VisualScore. 238 college students were recruited and used an E-book system to reading a learning material. The learning reading behavior consists of 11 basic reading variables, such as PC, Mobile, Tablet, Bookmake, Memo, Highlight, Underline, Prev, Next, Readtime, and Readpage. We used three groups, including Information Gain (IG) and Gain Ratio (GR), ANOVA and χ^2, and Fast Correlation Based Filter (FCBF), to perform correlation analysis. In addition, the principal component analysis (PCA) was used for exploring the learning patterns behind learning styles.

3. Results

3.1 Model Performance and Feature Importance

The prediction performance of the six prediction models is based on the learning style data in Accuracy, Precision, Recall, F1-score, and Auc five metrics. On the whole, the DT model has an average score above 0.7. In contrast, the prediction performance of the other five models is unevenly distributed across the five indicators, and the scores fluctuate relatively widely.

In terms of feature importance, it is found that there is a nearly similar proportion in the contribution of features to prediction performance. Although 5 models except DT do not meet the basic requirement of good prediction performance, the common thing is that the VisualScore feature offers the most important contribution for prediction with the largest proportion (0.752 for DT, 0.240 for RF, 0.240 for XG, -0.459 for SVM, -0.046 for LR, 0.159 for KNN).

3.2 Correlation between Learning Style and Reading Learning Behavior

The study analyzed the relationship in terms of the amount of information, sample variability, and inter-sample distance respectively. It was found that the first correlation exists between the Prev (IF 0.023, GR 0.011) behavior, which refer to scrolling back to read the material, and the VisualScore learning style. Second, Memo (ANOVA 8.501, χ^2 4.497) and HighLight (ANOVA 5.909, χ^2 3.199), also called mark behavior, are highly correlated with learning style. Specifically, from the IG, GR, and FCBF indicators, the Prev feature occupies first place, with scores of 0.023, 0.0115, and 0.017 respectively.

3.3 Learning Patterns behind Learning Style

The study used principal component analysis to extract common factors on students' reading behavior and then organized them into mutually independent categories. First, for the analysis of learning patterns of students with visual learning tendency, their reading learning behaviors were analyzed by PCA, and the 11 features were extracted as principal components to form four common factors, which are time-spending category (Next 0.94, Readtime 0.927, and Readpage 0.9), nark category (Highlight 0.889, Memo0.784 and UnderLine 0.631), repeated reading category (Prev 0.659 and PC 0.665), and mobile device category (Tablet 0.779 and UnderLine 0.297)

Second, the learning pattern of students with verbal learning tendencies was determined. There are three categories of learning patterns. The first one is a repeated reading category (Prev 0.951, Readpage 0.913, and Readtime 0.819), followed by the second category, mark category (Memo 0.933 and Highlight 0.868), finally, mobile device category (Mobile 0.633 and Underline 0.441).

4. Conclusion

Prediction model selection and feature importance. In terms of binary classifications that based on learning style data, DT model performs best. For the prediction performance, it was found that the DT model outperforms other models, with an average score of above 0.7. This result not only provides an insight of impacts and contribution of learning style data to predict students' achievement, but also figures out the extent to which various categories of learning styles affect the students' achievement, as well as ranks the importance of learning styles.

Regarding the feature importance, it is obvious that the VisualScore feature contributes most to the model construction, no matter what kind of models are based on various algorithms. However, the largest proportion of contribution exhibited by the VisualScore feature occurs in the DT model. There are two points worth noting when performing feature importance calculations. First, the feature importance calculation uses different calculation methods, and this diversity phenomenon is determined by the algorithm behind the model. Second, in terms of feature importance some models are limited by the characteristics of the algorithm behind them, and tend to equalize the feature contributions during model crating, such as the KNN model.

The results obtained from multiple correlation calculation methods are better than the analysis results under a single dimension. This study used several different correlation analyses between learning styles and reading behaviors. The analysis revealed that the two behaviors, repeated reading and marking, are highly correlated with learning styles. The former was explored from the perspective of informativeness, concentrating on the informativeness and uniqueness of the behavior to the learning style representation. The latter puts the analytical perspective on the variability of the sample group and within-group, emphasizing the variability between learning styles and reading behaviors.

Learning patterns behind learning style. For students who tend to be visual or verbal, repeated reading and marking learning methods are the most important learning mode. Based on the results of the PCA analysis, it was found that, on the one hand, verbal and visual learning styles have similar learning patterns, such as repeated reading, marking, and using mobile devices. On the other hand, they show differences the time-spending. In particular, students with visual-prone learning styles show the time-spending pattern. However, this difference has not been further confirmed and there is a lack of understanding of why it occurs, which will be a focus of research in the future.

Acknowledgements

Part of this research work was supported by the Grants-in-Aid for Scientific Research Nos. 21H00905 and 18H04125 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan. We would like to thank all the people who prepared and revised previous versions of this document.

References

Hames, E., & Baker, M. (2015). A study of the relationship between learning styles and cognitive abilities in engineering students. *European Journal of Engineering Education*, *40*(2), 167-185.

Huang, A. Y., Lu, O. H., Huang, J. C., Yin, C. J., & Yang, S. J. (2020). Predicting students' academic performance by using educational big data and learning analytics: evaluation of classification methods and learning logs. *Interactive Learning Environments*, *28*(2), 206-230.

Yin, C., & Hwang, G. J. (2018). Roles and strategies of learning analytics in the e-publication era. *Knowledge Management & E-Learning: An International Journal*, *10*(4), 455-468.

Yin, C., Yamada, M., Oi, M., Shimada, A., Okubo, F., Kojima, K., & Ogata, H. (2019). Exploring the relationships between reading behavior patterns and learning outcomes based on log data from e-books: A human Factor Approach. *International Journal of Human–Computer Interaction*, *35*(4-5), 313-322.

Zhao, F., Hwang, G. J., & Yin, C. (2021). A Result Confirmation-based Learning Behavior Analysis Framework for Exploring the Hidden Reasons behind Patterns and Strategies. *Educational Technology & Society*, *24*(1), 138-151.