게임 기반 교육에서의 사용자 잔존 요인 분석

Analyzing key features affecting user retention in game-based learning

Yu Sik Kim○ June Young Park Kee Jun Han Mun Yi
Korea Advanced Institute of Science and Technology

Abstract

In this study, we investigate crucial factors affecting engagement in game-based learning environment. As a measure, we used different classification algorithms including stacking model to classify user retention as early as possible in the game and analyzed important features selected by the classifiers. Having prior experience in digital games, plays significant role when classifying user retention in early stages. Additionally, we suggest balancing learning time and game time for effective design to further endorse continued learning.

1. Introduction

Game-based learning is a term used for incorporating learning materials in games to enhance engagement while challenging with learning contents. Recent studies have found that game-based learning have positive effects on academic performances while keeping high engagement [1][4]. With high engagement, game based learning generally increases user motivation to learn challenging contents [4].

A method to measure engagement and motivation to learn in game-based learning is to find user retention rate or dropout rates. Accurate classification of user retention require important determinants, which can inform us about game designs and its contents. Digital game industries focused on detecting churn users as early in the game as possible in order to maximize profits. Many studies in line have already discovered certain behavioral patterns that affects user retention in gaming industries [3]. However, in game-based learning, many studies assume that engagement and motivation come with the game part [2]. According to Schimanke, addictive games make users spend time with the same content repetitively but in educational games if users spend time repetitively with the contents that they already know, they could be very counterproductive and thus lose engagement [6]. Therefore, analyzing factors affecting the user engagement to increase retention rate should be highlighted in game-based learning as well.

Our key findings in this study are that users with prior experience, which is called mastery level in this study, in digital games positively affects user retention rate, which implies higher academic performance and game time to learning time ratio suggest repetition effects may play key role in game-based learning.

2. Classification models

In this section, we present classification models used for the experiments. As a baseline, we compared results of popular classification models, which are Random Forests (RF), Logistic Regression (LR) and Support Vector Machines (SVM). To prevent from overfitting and under fitting, we also adapted a type of ensemble method called stacking model. Stacking model combines results from several classification models from the first layer and uses it as input to the second layer to calculate the final prediction.

3. Data and pre-processing

We developed a mobile game for studying English vocabularies as shown in Figure 1. Figure 1(a) is the start of a session, followed by study (b) and quiz scenes (c). The session ends with the game showing the leader board at the end (d). We collected user log data and pre-survey data from 64 participants over 3 day period with base retention rate of 35.9%. According to the data, the average number of sessions is nineteen, which we chose to be the decision point of selecting return users and churn users in this binary classification task. As many studies have noted, a successful game-based learning applications highly engage users while showing improvement on academic performance [6]. As shown in Table 1, return users have higher quiz score on average, which validates that this is a successful system. Furthermore, many studies include game play features and behavioral features in churn prediction problems [3]. On top of those features, we added educational features and personal traits obtained from pre-survey for this study as shown in Table 2.

Table 1. Basic statistics of master data.

	Churn	Return
Avg_quiz_score	80.9 %	82.9 %
Avg_game_score	3294.11	5442.86
Session interval	267.74 min	190.41 min
Avg_study_time	12.09 min	2.09 min
Avg_quiz_time	1.96 min	1.25 min
Avg_game_time	30.11 s	38.95 s
Game:Learn ratio	0.2	0.5
Avg_num_sessions	13.27	36.45
Num_users	39	25

Table 2. Terms and defintions of used features.

Feature type	Feature name	Description
Game Play	Game time	Time spent playing game per session
	Game score	Game score per session
Educational	Study time	Time spent studying vocabulary per session
	Quiz time	Time spent solving vocabulary quiz per session
	Quiz score	Quiz score per session
	Learning time	Sum of study time and quiz time
Behavioral	Session interval	Time between each session
	Game to Learn ratio	Ratio between game time and learing time
Subjective	Mastery	Self-reported prior game experience

4. Experiment method

We selected popular classification models such as Random Forest (RF), Logistic Regression (LR) and Support Vector Machines (SVM) as well as a type of ensemble model called stacking. Hyper parameters for all the models are tuned using grid search for the best classification accuracy. Then we performed 10-fold cross validation to compare accuracy score, precision/recall and f1-score as shown in Table 3.

As previously mentioned, 35.9% of users returned implying that naïve baseline accuracy of random prediction is 64.1%. Each model carried out feature selections for the best output. While the normal stacking model performed the best in all cases, our key finding is achieved with adding self-reported prior game experience feature from pre-survey, which boosted the stacking model accuracy to 95%.

Figure 1. Screenshots of English Vocabulary game.

6. Analysis

6.1 Model comparison

We ran all the models using data from first session to first seven sessions because we assumed that discovering important features that affect the user retention early in the game is crucial in game designs. According to Table 3, the classification accuracy score resulted below the naïve baseline accuracy when using only the first session data. As the data size increased, the accuracy score also improved for stacking model and random forest. Overall, the stacking model performed the best with accuracy score 0.7576. Using the best resulted model comes the main finding of this study. When we included mastery level feature in the dataset, stacking model performed much better in all cases. Using the 5 session datasets, the model scored 0.95 in accuracy and 0.96 for precision/recall and f1-score. For the first 7 sessions of data, the model only improved slightly.

6.2 Feature importance

Analyzing the relationships between independent features and retention variable helps to inform important factors in determining user retention [3]. To do so, we examined feature importance of the stacking model with pre-survey data as shown in Table 4. Although mastery feature is self-reported from the pre-survey, it is the most important factor in this study with Gini importance value 0.1835. This suggests that the prior experience and exposure to digital games cannot be underestimated. Ranked in second with Gini importance value 0.1532, session interval plays important role as well. Shorter session interval is directly related with number of sessions played, which implies higher retention rate in this study. Finding an optimal session interval is one of the widely researched areas in digital games and game-based learning [3][6]. Interesting features are the time related features. Game time to learning time ratio implies the balance of entertainment

Table 3. Comparison of different classifier models.

Number of sessions	Classifier	Accuracy	Precision	Recall	F1-score
1	RF	0.619	0.68	0.62	0.63
	LR	0.6190	0.62	0.62	0.62
	SVM	0.667	0.44	0.67	0.53
	Stacking model	0.5714	0.62	0.57	0.58
	Stacking model with pre-survey	0.6190	0.65	0.62	0.63
3	RF	0.7049	0.72	0.70	0.71
	LR	0.6229	0.67	0.62	0.64
	SVM	0.7213	0.52	0.72	0.60
	Stacking model	0.7213	0.74	0.72	0.73
	Stacking model with pre-survey	0.8536	0.86	0.85	0.86
5	RF	0.7171	0.71	0.72	0.70
	LR	0.6767	0.67	0.68	0.67
	SVM	0.626	0.39	0.63	0.48
	Stacking model	0.7576	0.76	0.76	0.74
	Stacking model with pre-survey	0.95	0.96	0.96	0.96
7	RF	0.7686	0.77	0.77	0.76
	LR	0.6641	0.67	0.66	0.66
	SVM	0.6268	0.39	0.63	0.48
	Stacking model	0.7761	0.77	0.78	0.77
	Stacking model with pre-survey	0.96	0.96	0.96	0.966

and learning. According to Table 1, the returned users have higher game to learn ratio. In other words, returned users tend to spend more time playing games than churned users. This is interesting factor to further investigate in future studies.

Table 4. Feature importance

Features	Gini importance
Mastery level	0.183528
Session_interval	0.153201
Quiz time	0.146726
Study time	0.095776
Game : Learn ratio	0.094568
Game time	0.080133
Learning time	0.069765

7. Conclusion and future works

Substantial amount of studies have been published addressing classification of user retention in digital game industries [3] and game-based learning [6]. In this study, we adapted stacking model and compared with popular classification algorithms such as Random Forest, Logistic Regression and Support Vector Machines. Stacking model outperformed the rest of the models in classification accuracy. From the stacking model, we were able to discover significant features such as mastery level, session interval and learning time and game time ratio. Based on these findings, we will further develop our game to study the influence of balancing learning time and game time, which may have positive effects on user retention.

Works Cited

[1] All, A., Castellar, E., Looy, Jan.: Assessing the effectiveness of digital game-based learning: Best practices. Journal of computers & education, 92-93, pp. 90-103, (2016).

[2] Denny, P., McDonald, Fiona., Empson, R., Kelly, P., Peterson, A.: Empirical support for a causal relationship between gamification and learning outcomes. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM, New York, NY, USA, Paper 311, 13 pages. (2018)

[3] Drachen, Anders., Lundquist, E., Kung, Y., Rao, P., Sifa, R., Runge, J., Klabjan, D.: Rapid prediction of player retention in free-to-play mobile games. In proc. The Twelfth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-16), pp. 23-29, (2016).

[4] Hamari, J., Koivisto, Jonna., Sarsa, Harri.: Does Gamification Work? – A literature review of empirical studies on gamification. 2014 47th Hawaii International Conference on System Sciences, Waikoloa, Hi, pp. 3025-3034, (2014).

[5] Hamari, J., Shernoff, D., Rowe, Elizabeth., Coller, B., Asbell-Clarke, J., Edwards, T.: Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. Computers in Human Behavior 54, pp. 170-179, (2016).

[6] Schimanke, F., Mertens, R., Vornberger, O.: Designing for Motivation: Design Considerations for Spaced-Repetition-Based Learning Games on Mobile Devices. E-Learn 2014, New Orleans, LA, USA, (2014).