Althea Vail Wallop

CS 229 Final Project Report

Learning to Predict Disengagement in Educational iPad Application

Abstract

The focus of this project is to evaluate game design in an educational iPad application designed for young children and to build a predictive model for learner disengagement. I apply machine-learning techniques to learner activity data collected from an iPad application called Leo's Pad to predict likely learner disengagement based on gameplay frequency and performance.

Background and Purpose

Research on cognitive development has revealed the importance of early childhood education as a predictor for later academic success. Learning in young children is highly dependent upon engagement, since attention span is often a limiting factor for preschool aged children. Consequently, curricula for kindergarten readiness targeted toward children aged 3 to 5 years would benefit greatly from incorporating game elements including animations and interactive play. Computerized, animated games that may be more engaging and easy to use touch-based technology provide an opportunity for advancement toward interactive early childhood education (Chion & Schuler, 2010; Melhuish & Falloon, 2010; Carlson, Moses & Breton 2002; Bialystok & Senman, 2004).

Leo's Pad is an iPad application made by Kidaptive, Inc., which consists of an animated series of "appisodes" containing a balance of cinematic content and interactive games aimed at supporting metacognitive development in preschool learners. Additionally, the application uses embedded assessments to customize learning. Ipad applications like Leo's Pad may be able to leverage young children's interest in interactive games and animation, combined with opportunity for customized teaching, to optimize development along certain cognitive dimensions such as working memory and cognitive inhibition.

Given the ability computerized applications have for capturing fine-grained measurements of learner activity, and since engagement is a necessary factor for learning, a predictive model of learner engagement given prior gameplay history seems a suitable first step to making individual learner recommendations and better understanding which game elements foster engagement in young learners. This project sets out to build a predictive model of learner disengagement and to examine which features of learner activity indicate engagement. These predictions may lend further insight into what kinds of game elements are particularly salient for preschool learners.

Data

The back-end database of Leo's Pad captures learner activity data including gameplay events and milestone performance. Gameplay events are registered whenever a learner plays a game (not simply when they log onto the app). "Milestones" are either granted or revoked depending on how well the learner played the game. Milestone performance measures occur at completion of each game and can be thought of as "levels" of a game that also reflect the player's performance. For this project, I dissected frequency of gameplay events from this database activity data. I tallied the gameplay frequency per week for each learner and then discretized these tallies into the following buckets: 0-10, 11-20, 21-30, 31-40, and >50. I also parsed milestone events and created features representing grants (level-ups) and revokes (level-downs) for each learner per week. I used weekly gameplay and milestone scores for each learner for three weeks as the features for my learning algorithms. The learners were then labeled as "disengaged" (y=1) and "engaged" (y=0) based on whether they logged any gameplay activity during the three-week disengagement period following the input data. The sample training set used for this project consisted of gameplay events for more than 10,000 individual learners collected during six weeks.

Methodology

I framed my prediction of learner disengagement as a classification problem using supervised learning given input features that contain gameplay frequency and performance data. Since the data is not Gaussian, I chose following classification algorithms: Naive Bayes classifier with Laplace smoothing, SVM, and logistic regression. The data set was split into training (70%) and testing (30%) sets, since the data set is sufficiently large to allow for this procedure for measuring generalization error. The data was normalized to include approximately equal data for both labels.

Results

For each classification algorithm, the training examples were determined by discretized gameplay frequency and performance as mentioned above. I tried the algorithms with three different features sets: frequency only, performance only, and frequency and performance together. All classification algorithms produced similar testing and training accuracy (see Table 1). These accuracies represent the best feature set out of all my attempts: a discretized version of gameplay frequency combined with separate features for milestone achievement for each of the three weeks prior to disengagement. The same algorithms using only features from one week prior to disengagement were less effective.

In the beginning of the term, I used a data set that had many learners who logged no gameplay during the time period of interest and thus had zeros in the feature columns. Naturally, each of these learners was considered "disengaged." This skewed the classification algorithms, especially SVM such that they were highly "accurate" (~99%) in performance on the training and testing set. However, the predictions were flawed because they were not actually good predictors using gameplay features since so much of the data had zeros predicting zeros. I corrected this error by examining only learners who played at some point during the 3 weeks prior to the disengagement period. Although this new dataset produced lower "accuracy" using the same classification algorithms, it proved to be a better predictor at classifying learners who had logged gameplay data and then disengaged during the disengagement period. Similarly, I noticed that the SVM produces more accurate testing results (~.85) when it focuses on learners who played during each of the 3 weeks prior to the disengagement period (had no zero features). However, these results were misleading again since this limited the training examples to a very small set and they were mostly disengagement labels. Since the purpose of this project was to predict learner disengagement based on their gameplay data, I chose to report the results with lower accuracy because I believe they are more applicable to solving the central issue.

Conclusions

SVM and logistic regression both produced relatively equivalent training and testing error. The closeness between the training and testing errors indicates a high bias in the data. The Naïve Bayes classifier had higher training accuracy but lower testing accuracy, indicating high variance in the data. The high bias and variance may be a result of random noise in the data. One possible explanation for this random noise is that it is difficult to determine whether reasons for disengagement are even game-related. For example, since preschoolers do not presumably have control over their own schedule, they may want to continue playing the game, but are forced to disengage because their parents need to use the iPad or they need to leave for preschool, etc. Another explanation is that engagement may be based on personal preference that cannot be captured in gameplay activity. Finally, because the data was based on back-end game event collection and not on structured experimental collection, it cannot be easily verified that all events represent the expected event of a preschool learner playing independently. Many preschoolers may play with the assistance of their parents or siblings or may have little control over which games they choose to play and when.

Although the classification algorithms produced somewhat reasonable accuracy, another trend emerged as a feature in the data: time. Figure 1 shows the disengagement rate of learners across the three weeks leading up to the disengagement period. The "fourth week" in red represents the three-week disengagement period. One interpretation of this data is that learners tend to disengage over time in a predictable manner. However, in this data set, we have no context of the learner's gameplay history outside of this six-week period. It may be the case that learners stay engaged for many weeks before our "week 1" marker. It may also be the case that the majority of our "week 1" learners were new to Leo's Pad and disengaged immediately. Yet another extrapolation could be that learners engage for a week, slowly disengage over the following three weeks (as the data indicates), but then re-engage following some event such as the release of a new game, or another visit to grandma's house and thus another chance to play grandma's iPad.

Implications for further research

As mentioned in the introduction section, this project is only meant as a preliminary pass at understanding how learners engage with a game-based application. Following the discussion in the previous section regarding timing of disengagement, the next steps of this project would be to look at the learner's entire gameplay history and how their disengagement maps over time. Furthermore, we may be interested in looking at how average learner engagement trends change centered around the release of a new "appisode" (new bundle of games) or during holidays (when children may have more time to play games with their parents). Looking at these trends was not within the scope of this project due to the laborious process required to collect and parse the data.

Another avenue for further exploration would be to use unsupervised learning techniques to cluster gameplay activity around certain games. This may lend insight into which games are most successful at retaining learner engagement. Clustering the most and least popular games by game elements may also allow us to see what particular game elements foster the most and least engagement.

In conclusion, this project provides evidence that we can indeed predict learner disengagement by using prior gameplay history whose features involve frequency of play and performance. Further exploration will involve digging deeper into which specific game elements may be involved in predicting learner disengagement.

Table 1:

	SVM	Logistic Regression	Naïve Bayes
Test Accuracy	.710	.6903	.6504
Training Accuracy	.703	.6961	.801

Figure 1:

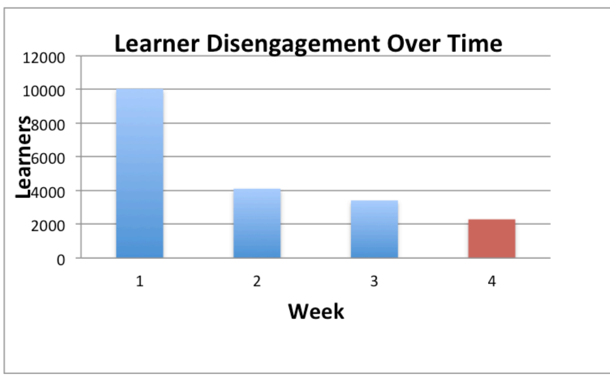
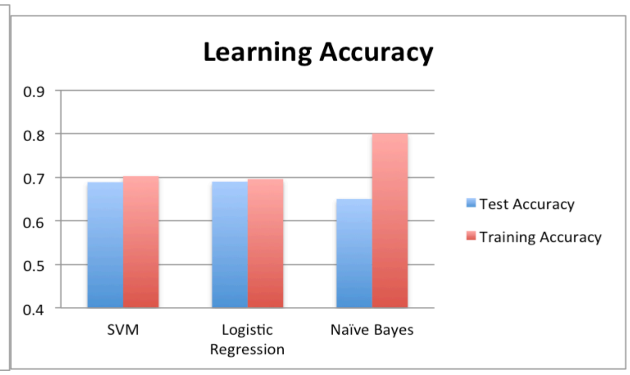


Figure 2:



Acknowledgements

The data set for this project was collected from back-end server data from Leo's Pad, made by Kidaptive Inc. David Hatfield contributed guidance in the planning process of this project by providing the dataset and extensively discussing the goals and interests of the project.

References

Chion, C. & Shuler, C. (2010). *Learning: Is there an app for that? Investigation of young children's usage and learning with mobile devices and apps.* New York: The Joan Ganz Gooney Center at Sesame Workshop.

Melhuish, K. & Falloon, G. (2010). Looking to the future: M-learning with the iPad. *Computers in New Zealand Schools: Learning, Leading, Technology, 22*(3).

Carlson, S., Moses, L., & Breton, C. (2002). How specific is the relation between executive function and theories of mind? Contributions of inhibitory control and working memory. *Infant and Child Development, 11*, 72-92.

Bialystok, E. & Senman, L. (2004). Executive processes in appearance -- reality tasks: The role of inhibition of attention and symbolic representation. *Child Development, 75*(2), 562-579.