Churn Prediction for High-Value Players in Casual Social Games

Julian Runge and Peng Gao
Team Sparkle
Wooga GmbH
Berlin, Germany
julian.runge@gmail.com
petergao0528@gmail.com

Florent Garcin and Boi Faltings
Artificial Intelligence Laboratory
École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland
florent.garcin@epfl.ch
boi.faltings@epfl.ch

***Abstract* — Predicting when players will leave a game creates a unique opportunity to increase players' lifetime and revenue contribution. Players can be incentivized to stay, strategically cross-linked to other games in the company's portfolio or, as a last resort, be passed on to other companies through in-game advertisement. This paper focuses on predicting churn for high-value players of casual social games and attempts to assess the business impact that can be derived from a predictive churn model. We compare the prediction performance of four common classification algorithms over two casual social games, each with millions of players. Furthermore, we implement a hidden Markov model to explicitly address temporal dynamics. We find that a neural network achieves the best prediction performance in terms of area under curve (AUC). In addition, to assess the business value of churn prediction, we design and implement an A/B test on one of the games, using free in-game currency as an incentive to retain players. Test results indicate that contacting players shortly before the predicted churn event substantially improves the effectiveness of communication with players. They further show that giving out free in-game currency does not significantly impact the churn rate or monetization of players. This suggests that players can only be retained by remarkably changing their gameplay experience ahead of the churn event and that cross-linking may be the more effective measure to deal with churning players.**

Keywords* — *churn prediction; neural networks; hidden Markov model; A/B evaluation; social casual games; freemium

I. Introduction

A social game is a type of online game that facilitates player interaction through social networks, and typically features multiplayer and asynchronous gameplay mechanics. Social games were traditionally implemented as browser games, but have recently seen drastic growth on mobile devices. The social gaming industry has been experiencing fast-paced growth over the past years with estimated revenue of 5.4 billion Euros in 2012 and expected revenue of 10.7 billion Euros in 2016. By 2016, social games are expected to account for nearly 50% of the video game market [1]. Casual games are characterized by shorter session length, simpler rules and lower required commitment in comparison to traditional core games.

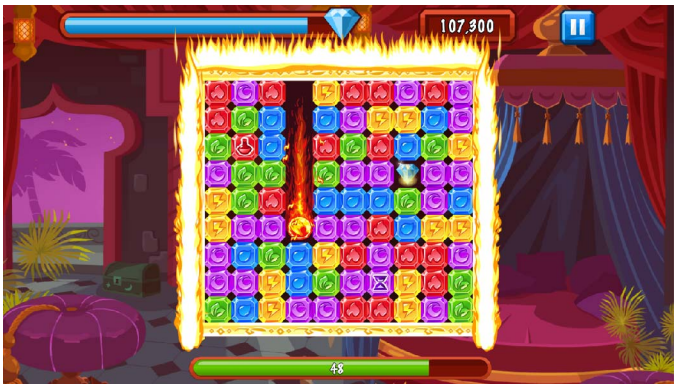

Figure 1. Screenshot of Diamond Dash iOS

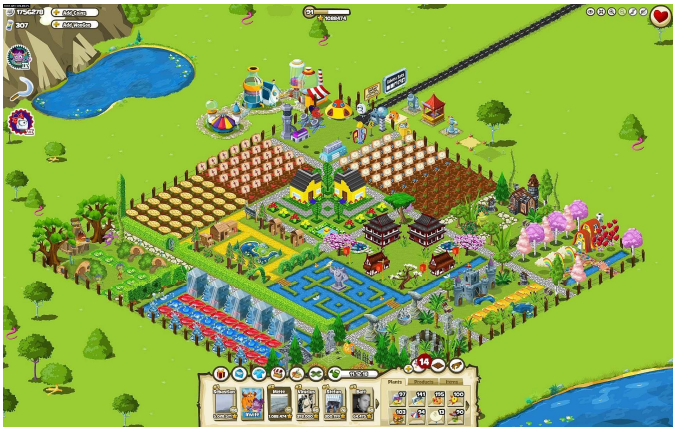

Figure 2. Screenshot of Monster World Flash

Most social casual games make use of the freemium business model. Freemium – or free-to-play – means that the game is offered to players for free and they can purchase upgrades and special items, unlock additional features or increase progression speed by spending money on in-app purchases. Offering free entry into a game vastly facilitates player acquisition, but only a few players ever spend real money on in-app purchases – far below 10% for most social casual games. The freemium model creates a non-contractual relationship with the customer. Leaving the game is very easy for players. One of the major goals of a social game developer

978-1-4799-3547-5/14/1.00 ©2014 IEEE

is to extend the lifetime of a player in the game to increase the player base and potential for monetization. The ability to predict when a user will leave a game opens up an opportunity to adjust the game-playing experience to extend the lifetime of a user in a game or to ignite a new lifetime in another game. To achieve this, players can be incentivized to stick with a game, cross-linked to another game in the company's portfolio or even cross-sold to other companies. In this paper we design, implement and evaluate a churn prediction model for high-value players based on players' gaming activity data in two very large social games by Wooga and test incentivization as a method to deal with churning high-value players in one of the games.

The two live games chosen for the study are Diamond Dash iOS and Monster World Flash. Figure 1 and Figure 2 show screenshots of the two games. Diamond Dash is a block-puzzle game, like Tetris or Bejeweled, on mobile devices. The player has to clear diamonds of the same color as quickly as possible in a specified time in order to beat the high score. Cleared diamonds are replaced from the top. There are also several power-ups or boosts present in the game that can make clearing the diamonds faster, so players can obtain a higher score and also have a more exciting gaming experience. Monster World is a farming game on Facebook which simulates the economics of running a highly personalized garden. The player plants and harvests crops to obtain coins and other in-game currency. As the players climb through various levels, they have the opportunity to extend and beautify their garden, and unlock new game features. Unlike Diamond Dash, Monster World is driven by missions that the players are asked to complete. This requires more regular interaction and engagement from players.

In this paper, we define the high-value player segment and the churn event for the two games under study. Further, we formulate churn prediction as a binary classification problem. We then compare the prediction performance of four different classification algorithms and attempt to explore the temporal dynamics of time series data using a hidden Markov model (HMM). In order to evaluate the business impact of a churn prediction model, we design and implement an A/B evaluation over one of the two live games.

II. LITERATURE REVIEW

Extensive research efforts [2, 3, 4, 5, 6] have been carried out in various domains to study customer churn and explore different modeling techniques to improve prediction performance.

Many machine-learning algorithms, such as decision trees, logistic regression and neural networks have been applied to churn prediction problems. [2] proposes a neural network based approach for predicting customer churn in cellular network services and claims that medium sized neural networks perform best for customer churn prediction when different neural network topologies are explored. [3] compares support vector machines, logistic regression and random forests over a newspaper subscription dataset and reports that the predictive capabilities of random forests outperform other algorithms for the dataset under study. A similar algorithm

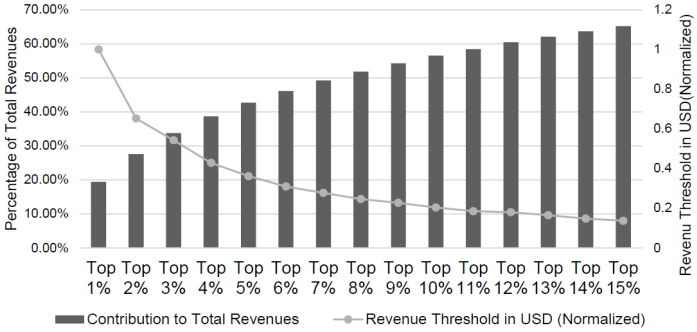

Figure 3. Contribution of Top Percentile Paying Players to Total Revenues in Diamond Dash

comparison study for a pay TV company includes a probability estimate of customer churn obtained from a Markov Chain model as an additional predictor in the churn prediction model and reports improved prediction performance [4]. [5] claims that neural networks are superior in prediction performance, as opposed to other models for churn prediction. Similar results can be found in [6] and [7].

Recent studies [8, 9] also try to approach churn prediction via social network mining or extending conventional machine-learning techniques with social network statistics. Social network analysis has also been used to predict customer churn in mobile networks [8]. The basic idea behind this is that the probability of a customer churning increases if his neighbors in the social graph have already churned. [9] constructs a hybrid prediction model extending the traditional tabular churn models with social network features and reports that traditional tabular churn models still score best.

To the best of our knowledge, there only is one other study [13] investigating churn prediction in free-to-play games. It was however unpublished when this research was conducted. Two relevant research works [10, 11] focus on churn prediction in massively multiplayer online role playing games. [10] proposes a social influence based approach that applies a modified diffusion model with different traditional classifiers. The social ties between players are characterized by the number of quests finished together by players. [11] proposes the three semantic dimensions of engagement, enthusiasm and persistence to construct derived features for churn prediction modeling and a hybrid classification approach based on weighted distance from labeled time series clusters. [12] takes a different approach and applies techniques from lifetime analysis on data from five First-Person Shooter games. [12] concludes that an average player's interest in the games evolves according to a non-homogenous Poisson process. Therefore, the data for initial playtime behavior of players can be used to predict when they stop playing.

Our main contributions are: To the best of our knowledge, together with [13] we are the first to thoroughly investigate the feasibility of churn prediction in free-to-play casual social games. Second, we are the first to assess the business impact – the effect on communication with players, churn rate and revenues – that can be derived from high-value player churn prediction for a big casual social game. We thereby discover several issues from a business perspective that cannot be found

from a pure data analysis point of view. Third, we explore the value of adding HMM features to a conventional classification model for churn prediction.

III. PROBLEM DEFINITION

A. Defining High-value Players

The term "high-value players" is a rather vague term. In order to pin down a more precise definition for high-value players, we looked into the contribution of top percentile paying players to the total revenue for Diamond Dash as shown in Figure 3. The top 7% of paying players contribute around 50% of the total revenue. The minimum revenue threshold (the minimum revenue generated by a player in order to make it to the top percentile) starts to flatten out below the top 10% of paying players. Players become increasingly homogenous when further extending the percentile. The top 10% seem to catch all the players with an exceptionally high value which led us to adopt the following definition:

Definition 1. *A* high-value player *in the game on day t=0 (the observation time) is a player that ranks in the top 10% of all paying players sorted in decreasing order of revenue generated by each player between days t = -90 and t = -1.*

B. Defining Activity

Since we want to reach players before their actual churn event, the prediction should be targeting *active* high-value players. We define *active* high-value players as follows:

Definition 2. *An* active high-value player *on day t=0 is a high-value player who has played the game at least once between days t = -14 and t = -1.*

C. Defining Churn

The term "player churn" defines a player that has permanently left the game. This decision may be conscious or not, driven by external or internal reasons. In practice, we need a threshold value for days of inactivity that we can use to clearly define the churn of a player. We considered the distribution of days of inactivity between logins for high-value players for Diamond Dash. For example, if a high-value player played the game on day t = 1 and day t = 3, then again on day t = 7, this gives two samples of the days of inactivity: one is 3 - 1 - 1 = 1 and the other is 7 - 3 - 1 = 3. Figure 4 shows the histogram of the distribution and cumulative distribution curve of days of inactivity. It shows that less than 2% of high-value players stay away from the game for more than 14 days. Hence, 14 days of inactivity is a good indicator of churn. With this definition, 98% of the players defined as churners truly churn.

Definition 3. *An active high-value player is said to be* churning *on day t=0 if she starts a period of 14 consecutive days of inactivity on any of the days between day t=0 and day t=6.*

D. Problem Statement

We model the churn prediction problem as a binary classification task where the goal is to assign a label **churn** or

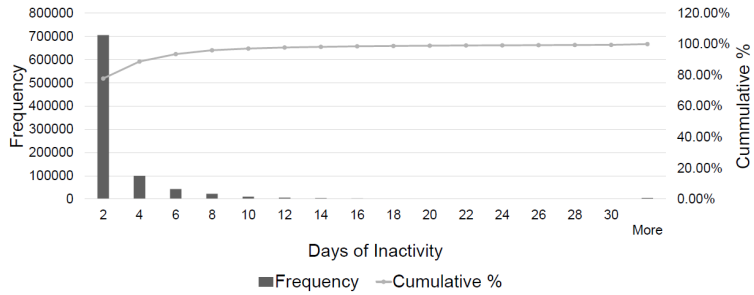

Figure 4. Histogram of Days of Inactivity for Diamond Dash

no churn to each player. We train various classifiers on labelled data of previously observed player behaviors up to a given day, and predict whether a player will churn or not within the week following that day. We use AUC, i.e. the area under the receiver operating characteristic (ROC) curve, for performance comparison because it allows us to compare models across all possible classification thresholds. ROC curves are commonly depicted in a chart with the false positive rate (FPR) on the x- and the true positive rate (TPR) on the y-axis. Classifier performance can be compared for different combinations of TPR and FPR and hence for different threshold choices. As AUC is the area under the ROC curve, when it is one, the classifier performs perfectly and the ROC curve follows the left and top border of the chart. Regardless of the threshold the TPR of the classification then is one and the FPR is zero. When the ROC curve is a diagonal from the lower left to the top right, AUC is 0.5 which reflects the case of random classification. Summarizing, our problem is to find the classifier that most correctly assigns the churn label to players across all possible classification thresholds and hence maximizes AUC.

IV. MODELING

A. Dataset

For both Diamond Dash and Monster World, we extracted the relevant historical tracking data of high-value players for two observation days, July 1st and August 1st 2013. We then constructed two labeled datasets to build the churn prediction model. Table 1 shows a summary of the raw dataset. There are three main categories of data: first, in-game activity tracking data such as a time series of logins per day or a time series of accuracy[1]; second, revenue-related tracking data, such as a time series of revenues generated by players; third, player profile data, such as how long the player has been playing the game and which country the player is from. We processed the data to alleviate the high positive skewness of the dataset by applying a Box-Cox transformation. Since the algorithms we used usually perform better with a standardized dataset, we further centered and scaled the data during the data preparation phase.

[1] Accuracy addresses successful taps, i.e. taps that removed sets of gems, divided by total taps. A set of at least three gems of the same color is needed for a successful tap.

Dataset	Diamond Dash	Monster World
# of players	10736	7709
# of churners	1821	352
# of non-churners	8915	7357
# of attributes	516	699

Table 1. Summary of Raw Dataset

Dataset	Final Feature Set
Diamond Dash	Time series of rounds played, accuracy, invites sent; days in game, last purchase, days since last purchase
Monster World	Time series of logins, level, in-game-currency 1 balance (WooGoo), currency 2 balance (Magic Wands)

Table 2. Final Feature Set

B. Feature Selection

The prepared datasets include more than several hundred attributes and not all of them are likely to be informative for making predictions. To identify the set of attributes to be used for the prediction, we performed a series of feature selection procedures. During feature selection, we used logistic regression with 10-fold cross validation to estimate the AUC performance of different feature sets. We experimented with the length of time series, eliminated time series that were highly correlated with others, and applied forward feature selection. Table 2 summarizes the feature sets for the final models. Empirical experiments showed that using only the last 14 days of historical data prior to the churn event yields the best prediction performance in terms of AUC.

C. Offline Evaluation

We compared the optimal prediction performance of neural networks (NN), logistic regression (Logistic), decision tree (DT) and SVM (support vector machine). For the SVM we used the radial basis function kernel and applied a parameter grid search to tune the hyper-parameters with 10-fold cross-validation. We experimented with 100 combinations of C (ranging from 0 to 5) and gamma (ranging from 0.2 to 5) with quadratic step size. The neural network has a simple one-hidden layer network topology. The number of hidden nodes is set to be equal to the sum of the number of attributes and classes divided by two plus one. Also for the neural network we used a parameter grid search with 10-fold cross-validation and tried 100 combinations of learning rate and momentum (both from 0.5^1 to 0.5^{10}).

Table 3 reports average AUC performance of different algorithms over the two datasets. Figure 5 and Figure 6 show the ROC curves for the four mentioned algorithms. The results are consistent over the two datasets. The performance of neural network and logistic regression is equal, but neural network is slightly better if we compare purely based on AUC. The performance of SVM and decision tree is less satisfying. SVM performs better for low FPRs and decision tree performs better for FPRs higher than 25%. For us, a low FPR is more important because we deem messing with non-churning players more harmful than neglecting some of the churners. Hence, the SVM would be preferred to the decision tree. Also in terms of AUC the SVM outperforms the decision tree.

Classifier	Diamond Dash	Monster World
neural network	0.815	0.930
logistic regression	0.814	0.924
decision tree	0.732	0.850
support vector machine	0.762	0.903

Table 3. AUC Performance of Different Algorithms

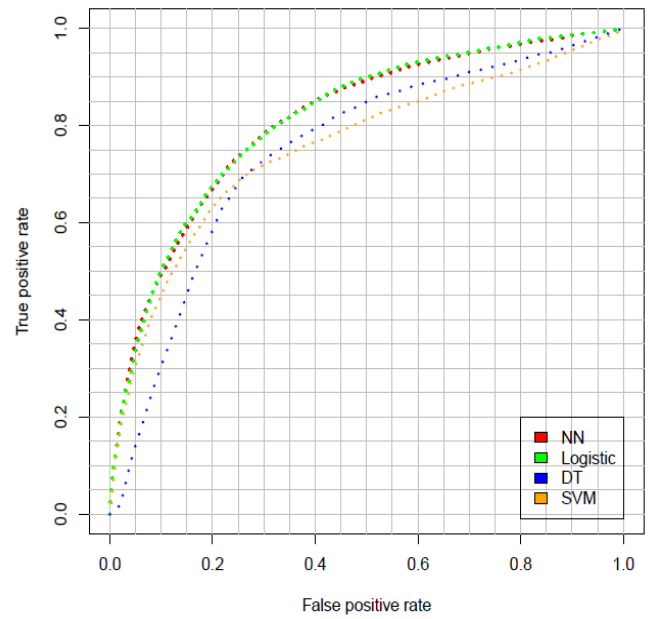

Figure 5. ROC Comparison on Diamond Dash Dataset

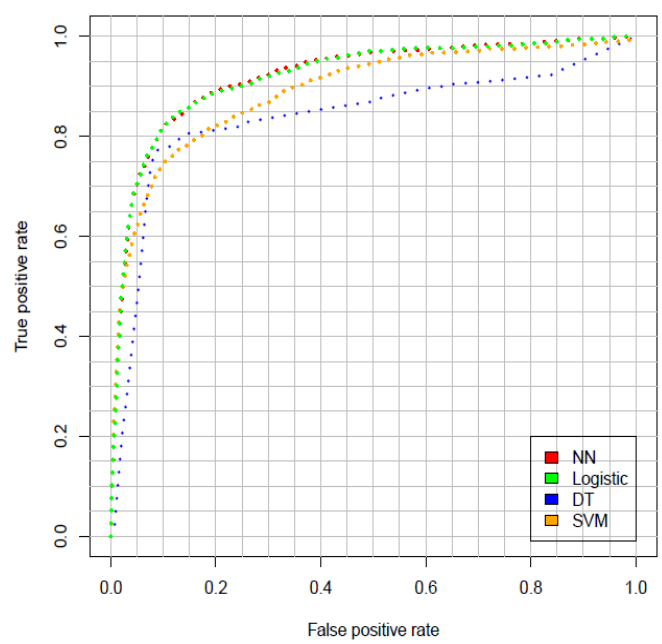

Figure 6. ROC Comparison on Monster World Dataset

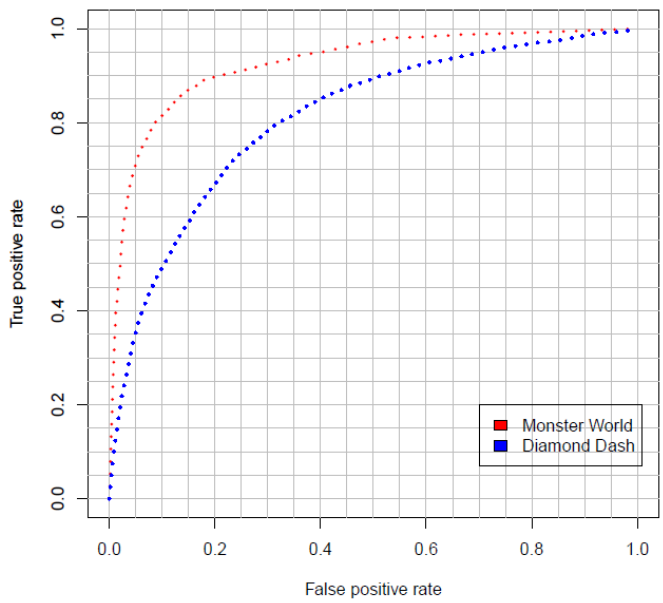

Figure 7. Comparing Diamond Dash and Monster World

D. Comparison of Prediction Performance

Figure 7 shows a comparison of the ROC curves for a neural network on Diamond Dash versus Monster World data. The same prediction modeling technique performs much better for Monster World than for Diamond Dash. More specifically, if we fix the FPR – that is the percentage of actually non-churning users we include in the predicted list of churns – at 5%, we achieve a TPR higher than 70% for Monster World. Hence, we reach more than 70% of truly churning users, while for Diamond Dash we only reach 35% of truly churning users for the same FPR.

The intuitive explanation of this difference is that the nature of the games differs. Monster World, though being a casual game, requires more constant interaction from the players and is more engaging. It has many additional features on top of the core farming mechanic. Among these are crafting and selling products, lotteries, an underwater garden and deep social features like visiting friends' gardens. Diamond Dash on the other hand fully focuses on timed rounds of the same core mechanic. It does not require a high level of commitment from the players and allows for much more casual interactions. Logins per user per day are substantially lower in Diamond Dash and times between sessions for one player can span several days. The behavioral patterns of in-game activities in Diamond Dash thus are less pronounced compared to Monster World. They appear to be more random and contain less structured information on an upcoming churn event.

E. Combining Neural Network and HMM

Though the modeling techniques discussed thus far already deliver fairly good prediction performance, one common issue with all the techniques considered is that they do not take the temporal dynamics of the time series attributes into consideration explicitly. If we switched the order of the data points in the time series, the resulting prediction would not be affected since (time wise) ordering of attributes is not accounted for by these algorithms. There are high quality historical tracking data dating back months and years available in Wooga's databases. In order to better leverage the information potentially present in these data, we turned our focus to HMMs. We included the results obtained from a HMM in the neural network to further improve the prediction performance.

We chose to use the Monster World dataset to model users' logins per day time series as a hidden Markov process because this dataset had previously proven to be better suited for churn prediction. The data under study are all instances of the logins per day time series for Monster World after the data cleansing step but without data transformation. This is because the data transformation alters the data in a way that makes them unusable for fitting a HMM.

The training data for the HMM can be denoted as a vector

$$\mathbf{L} = [L_1, L_2, \ldots, L_n]^T$$

where each $L_i = [L_i^{(-60)}, L_i^{(-59)}, \ldots, L_i^{(-1)}]$ is the time series of logins per day for a player between day t = -60 and t = -1 and n is the number of instances in the dataset.

We also made the following assumptions regarding the model:

- The instances of the logins time series are mutually independent, which is a valid assumption, since each instance is an observation of a certain different user.
- All instances of the logins time series are generated from one single underlying hidden Markov process. We hence assume that the HMM portrays an average user's stochastic behavior.
- The emission distribution of the HMM follows a Poisson distribution. Each value in the logins time series is a non-negative integer that records the number of login events.

Essentially the model setup reflects that the actual logins of a player on a certain date depend on the states of the players on that date. A state process which is hidden and unobservable is a Markov chain process. The actual observed values of logins follow a Poisson distribution, where the mean of the Poisson distribution depends on the state of a player on each date.

With the HMM, we leveraged more historical data and took the temporal dynamics into consideration. However, the HMM setup is hard to reconcile with our definition of churn and cannot be used directly for making predictions. In order to make use of the HMM that we built, we used it to extract new features to be added to the neural network. The idea is that this will enhance the prediction performance. We follow the approach of Burez and van den Poel [4], who incorporate features extracted from a Markov chain model into a Neural Network to improve prediction performance.

Through the HMM we calculated the following probabilities for each instance i in □ as new features to add to the neural network.

$$\mathbf{p_i} = [p_0, p_1, \ldots, p_{13}] \text{ where } p_k = \Pr(L_i^{(k)} = 0 | L_i = l_i)$$

Essentially, p_i is a vector with element p_k being the probability of player i not playing the game on date t=k given the observed sequence of L up to t = -1.

Under the setup of the HMM, the probability p_k can be easily calculated using the following equation.

$$p_k=(\alpha\mathbf{\Gamma}^{k+1}\mathbf{P}(0)\mathbf{1}')/(\alpha\mathbf{1}')$$

where α denotes the forward probabilities of HMM at t=-1, $\mathbf{\Gamma}$ is the transition matrix of the HMM, and $\mathbf{P}(0)$ is a diagonal matrix where the mth diagonal element is the probability of observing a 0, given the hidden state is m. Detailed mathematic proof of the above equation can be found in [14].

We then added the new features, $\mathbf{p_i}$, into the Monster World dataset after transformation and applied the neural network modeling on top of the new feature set. The AUC value of the model with the new feature set is 0.923, which degrades the prediction performance compared to the 0.930 we achieved with the old feature set. Inclusion of the HMM features appears to lead to overfitting which in turn reduces the AUC because of worse out-of-sample performance in the cross-validation. Only using the HMM features for prediction yields an AUC of 0.915. We hence did not include HMM results in our final proposed prediction model.

V. Business Impact

From the machine-learning perspective, we had built a satisfying churn prediction model. In reality, there will very often be business and/or technical issues that we cannot anticipate during the model building process. A way to figure out these issues is to integrate and deploy the prediction model on a small scale. In this section, we discuss the results we obtained by rolling out the prediction model to Monster World. We implemented an A/B evaluation to assess the business impact in terms of communication with players, churn rate and revenue. These dimensions define the strength of the model‘s use case.

A. A/B Test Set Up

The subjects under study were players of the game Monster World. All active players were randomly assigned to three test groups A, B and C. Groups A and B each included roughly 40% of the total player base and Group C included roughly 20% of the total player base. For Group A, the heuristic group, we sent out a substantial amount of free in-game currency, worth more than 10 USD, to high-value players after the churn event had occurred (14 days of inactivity). We call this the heuristic group because this measure can be taken without any predictive model in place. The gifts were sent both via Facebook application notifications and emails. For Group B, the predictive group, on top of the heuristic churn management policy of Group A, we used the churn prediction model to obtain a list of high-value players who are still active but about to churn, and sent incentives to this selective list of players as well. For Group C, the control group, no actions were taken to deal with churning high-value players.

The goal of this A/B test was to verify that the prediction model has actual business impact beyond what can be done without any predictive efforts. Business impact was assessed by means of key performance indicators (KPIs). The most important for the given use case were high-value player churn rate and revenue, as well as email campaign click-through rate and Facebook notification click-to-impression rate to measure effectiveness of communication with players.

The KPIs are defined as follows:

Definition 4. Churn Rate (CR)

CR = 1 – (# active high-value players) / (# high-value players)

Definition 5. Daily Revenues (DR)

DR = total revenues from players in a group during a day

Definition 6. Email Click Through Rate (CTR)

CTR = (# gifts claimed by email) / (# emails delivered)

Definition 7. Facebook Click To Impression Rate (CTI)

CTI = (# gifts claimed by notification) / (# notifications seen by players)

B. A/B Test Results

Table 4 shows the high-value player churn rate of different groups observed at the beginning and the end of the test. As can be seen, churn rates of high-value players increased strongly for all three test groups during the test. This is because Monster World is an old game (older than three years), now at the end of its lifecycle. Based on the assumption that player churn is a Bernoulli random variable we applied a Chi-square test to find out whether there are significant differences between groups. The results of the Chi-square test for churn rate are shown in Table 5. The null hypothesis is that the churn rate at the end of the test is equal for Group A and Group C and for Group B and Group C. Since both p-values are much greater than 0.05, there is no strong statistical evidence that the churn rate is positively affected by either of the two churn management policies. However, comparing the increase in churn rate between the start and the end of the test, we see a slightly slower increase in churn rate for Group B. This can be interpreted as indicative evidence for the higher effectiveness of a predictive churn management policy. We will outline later what could be the reasons for this higher effectiveness.

Historically for the game under study, the normalized daily total revenue follows a Gaussian distribution and has close to equal variance. Assuming this, we performed a two-sample t-test to test whether there is a significant difference in daily revenue between different test groups. The null hypotheses are that the means of the normalized daily revenues are equal between Group A and C and Group B and C. Results are shown in table 6. Since the p-values are again both substantially greater than 0.05, we are not able to reject the null hypothesis of equal means. Hence, there is no statistically significant evidence that the daily revenues are affected by the different churn management policies. Hence, there is no positive effect of the prediction model on revenues. These results are rather disappointing. However, for the last KPI of business impact the predictive model really made a difference. For Group A, the heuristic group, the CTR of the email campaign was 2.35%. For Group B, employing the predictive churn management policy, the same value was 11.90%.

	Group A	Group B	Group C
CR Start	7.80%	8.40%	6.77%
CR End	11.06%	11.39%	10.29%

Table 4. Churn Rate Results

	χ^2	p-value
A vs. C	0.3196	0.5718
B vs. C	0.667	0.4141

Table 5. Churn Rate Chi-square Tests Results

	T-statistic	p-value
A vs. C	1.0067	0.290853
B vs. C	0.77673	0.440707

Table 6. Revenues T-Test Results

	χ^2	p-value
A vs. B	4.5569	0.03279

Table 7. E-mail CTR Chi-square Test Results

	χ^2	p-value
A vs. B	24.1018	9.137e-07

Table 8. Facebook CTI Chi-square Test Result

Table 7 shows the Chi-square test results for comparing the email CTR between Group A and Group B. The p-value is less than 0.05. This indicates that with the prediction model we are improving the effectiveness of the email marketing and seizing the opportunity to reach out to the high-value players while they are still interested in the game.

We also further compared the funnel of the email campaign of Group A and Group B. Figure 8 illustrates the difference between the groups. At the top of the funnel we have the emails we sent out. On the next level, we lost about 10% of players for both groups in the delivery phase, since only about 90% of the email addresses recorded are valid. On the next level, 'Opened', we already see a clear difference in the opening rate. Finally, we have, on average, five times the click-through rate for the predictive group. And four times more players claimed the gift links when we used a proactive predictive churn management policy. Similarly, for Group A the Facebook notifications CTI was 8.66% while it was 31.48% for Group B. The p-value of a Chi-square test on the two groups is approximately 0 and therefore Group B performed statistically significantly better.

Figure 9 illustrates the comparison of the Facebook notifications funnel between Group A and Group B. The first step summarizes all players to whom we sent out notifications via Facebook. On the next level, there is already a difference in the delivery rate, since we are not able to deliver the notifications to players who have uninstalled the game. On the next level, 'Published', we further lost players. The reason for this is that if the player has blocked our notifications, Facebook will not publish any notifications to the players.

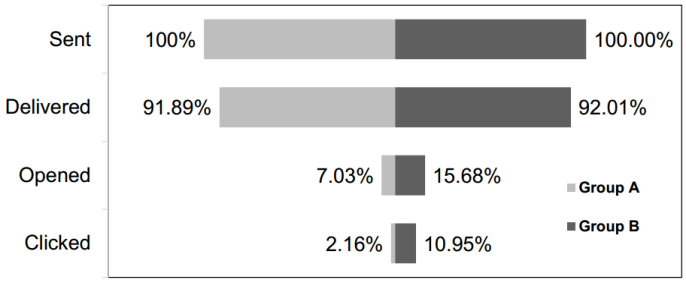

Figure 8. Email Funnel Comparison

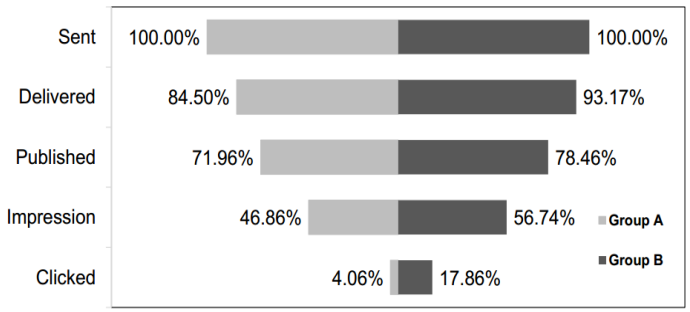

Figure 9. Facebook Notifications Funnel Comparison

For Group A and Group B, the percentage loss at this step is approximately equal. At the impression level, for Group A, only 46.86/71.96 = 65.12% published notifications are seen by players while for Group B, 56.74/78.46 = 72.31% published notifications are seen by players. In the final step, we witness a sizable improvement in the click rate of notifications. Notifications are three times more likely to be claimed in the predictive churn management policy Group, i.e. Group B.

A reasonable criticism of these results is that the good CTR achieved for the predictive group is driven by the amount of false positives in the group, i.e. players who are not about to churn, but are predicted to do so. When training and testing our algorithm, we consistently achieved a precision of better than 40% for repeated and out-of-sample testing. Assuming the same precision for the prediction used in the A/B evaluation, up to 60% of contacted high-value players may have been actual non-churners. In previous e-mail gift campaigns with high-value players for the same game we observed CTRs of around 10%. Assuming a similar CTR for the false positives among the predicted churners, the true positives – i.e. players who are about to churn – have approximately the same CTR, since the overall CTR is above 10%. So, contacting high-value players shortly before churn appears to be as efficient as contacting them while they are fully active. And much more efficient than contacting them after their churn event, where we observe a CTR of around 2%. The value of churn prediction here, therefore, is in enabling us to contact players at the very end of their lifetime in the game. Successfully crosslinking a player in the middle of their lifetime may be hurtful for the sending game because it is likely to shorten players' lifetimes in this game. On the other hand, crosslinking players who are at the end of their lifetime, and would leave the sending game anyway, comes at virtually no cost.

VI. CONCLUSIONS

In this paper, we provided a quantitative definition of the high-value player segment, defined the churn event and formulated churn prediction as a binary classification problem. We extracted relevant in-game activity tracking data, performed a series of data preprocessing procedures and obtained two high-quality datasets. We chose Diamond Dash and Monster World, two of Wooga's games with millions of players each, as subjects for our study due to their vast user base and long lifetime. We compared four different classification algorithms and attempted to include temporal dynamics in our classification through features extracted from a hidden Markov model. In order to evaluate the business impact of churn prediction, we designed and implemented an A/B test on Monster World, the game with substantially better prediction performance.

Experimentations showed that a single hidden layer neural network with fine-tuned learning rate and momentum outperforms other learning algorithms in terms of AUC. This is consistent across the two datasets we studied. For Diamond Dash we were able to achieve an AUC of 0.815 and for Monster World we were able to achieve an AUC as high as 0.930. Using the same modeling technique, the Monster World dataset responds better compared to the Diamond Dash dataset which suggests that churn prediction for highly casual games is more difficult. We see a very likely explanation for this in the fact that the behavioral patterns in highly casual games are less pronounced due to much less required commitment and engagement on the players' side.

A/B test results show that sending substantial amounts of free in-game currency (monetary value approximately 10 USD) to churning and churned high-value players does not affect the churn rate remarkably. This indicates that highly engaged players cannot be retained at the end of their lifetime by simple incentives. Compared to a reactive churn management policy, the one leveraging the prediction model improved the communication channel effectiveness by factor four to five. For email campaigns, the click-through-rate was increased from 2.4% to 12%. For Facebook notifications, the click-to-impression rate was increased from 8.7% to 31.5%.

A thorough assessment of the best strategy to deal with churning players in casual social games is highly valuable to industry and research alike. A possible extension of our research is an investigation of cross-linking to other (ideally similar) games in the company's portfolio. This appears to be a more promising approach to deal with player churn than incentivization. Another option that we discussed and that would be valuable to explore is to touch the deeper gameplay mechanics and change players' game experience in a way that keeps them interested in the game. Methodologically, the temporal dynamics preceding player churn could be scrutinized in more detail. Even though the integration of the HMM features failed to improve the prediction performance of the model, hidden Markov states may be a fruitful framework for actionable customer segmentation. Lastly, future research could beneficially address the differential predictive performance of algorithms for churn in browser versus mobile casual social games.

ACKNOWLEDGMENTS

We would like to extend our heartfelt gratitude to Wooga for granting access to relevant player data and allowing us to collaborate in such a pleasant and spirited environment. We also want to thank Jan Miczaika, Thomas Leclerc, Maciej Zak, Pia Simonsen, Robert Wolf, Markus Steinkamp, Armin Feistenauer and various other colleagues for their assistance in conducting our research.

REFERENCES

[1] iDate, Markets and Trends 2012-2016, market research report, [Retrieved from http://www.idate.org/en/Research-store/Social-Gaming 778.html on February 9th 2014].

[2] A. Sharma, and P. K. Panigrahi, "A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services," International Journal of Computer Applications, Vol. 27(11), pp. 975-8887, August 2011.

[3] K. Coussement, and D. Van den Poel, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques," Expert Systems with Applications, Vol. 34(1), pp. 313-327, January 2008.

[4] J. Burez, and D. Van den Poel, "CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services," Expert Systems with Applications, Vol. 32(2), pp. 277-288, 2007.

[5] T. Iwata, K. Saito, and T. Yamada, "Recommendation method for extending subscription periods," KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pp- 574-579, 2006.

[6] J. Hadden, A. Tiwari, R. Roy, and D. Ruta, "Churn prediction: Does technology matter," International Journal of Intelligent Technology, Vol. 1(1), pp. 104-110, 2006.

[7] J. Hadden, A. Tiwari, R. Roy, and D. Ruta, "Churn Prediction using complaints data," Enformatika, Vol. 13, 2006.

[8] K. Dasgupta, R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. A. Nanavati, and A. Joshi, "Social ties and their relevance to churn in mobile telecom networks," in EDBT '08: Proceedings of the 11th international conference on Extending database technology, (New York,NY, USA), pp. 668–677, ACM, 2008.

[9] P. Kusuma, D. Radosavljevic, F. W. Takes, and P. W. H. van der Putten, "Combining Customer Attributes and Social Network Mining for Prepaid Mobile Churn Prediction," Proceedings of the 22nd Belgian Dutch Conference on Machine Learning (Benelearn), pp. 50-58, 2013.

[10] J. Kawale, A. Pal, and J. Srivastava, "Churn prediction in MMORPGs: a Social Influence Based Approach," CSE '09: International Conference on Computational Science and Engineering, Vol. 4, pp. 423-428, August 2009.

[11] Z. H. Borbora, and J. Srivastava, "User behavior modeling approach for churn prediction in online games," International Conference on Social Computing (SocialCom) 2012, pp. 51-60, September 2012.

[12] C. Bauckhage, K. Kersting, and R. Sifa, "How Players Lose Interest in Playing a Game: An Empirical Study Based on Distributions of Total Playing Times," Proceedings of the Conference on Computational Intelligence and Games (CIG), 2012.

[13] F. Hadiji, R. Sifa, A. Drachen, C. Thurau, K. Kersting, and C. Bauckhage, "Predicting Player Churn in the Wild," Proceedings of the Conference on Computational Intelligence and Games (CIG), 2014.

[14] W. Zucchini, and I. L. MacDonald, Hidden markov models for time series: An Introduction Using R. Boca Raton, FL: Chapman & Hall/CRC, 2009.