Churn Prediction in Mobile Social Games: Towards a Complete Assessment Using Survival Ensembles

África Periáñez, Alain Saas, Anna Guitart and Colin Magne

Game Data Science Department

Silicon Studio

1-21-3 Ebisu Shibuya-ku, Tokyo, Japan

{africa.perianez, alain.saas, anna.guitart, colin}@siliconstudio.co.jp

Abstract—Reducing user attrition, i.e. churn, is a broad challenge faced by several industries. In mobile social games, decreasing churn is decisive to increase player retention and rise revenues. Churn prediction models allow to understand player loyalty and to anticipate when they will stop playing a game. Thanks to these predictions, several initiatives can be taken to retain those players who are more likely to churn.

Survival analysis focuses on predicting the time of occurrence of a certain event, churn in our case. Classical methods, like regressions, could be applied only when all players have left the game. The challenge arises for datasets with incomplete churning information for all players, as most of them still connect to the game. This is called a censored data problem and is in the nature of churn. Censoring is commonly dealt with survival analysis techniques, but due to the inflexibility of the survival statistical algorithms, the accuracy achieved is often poor. In contrast, novel ensemble learning techniques, increasingly popular in a variety of scientific fields, provide high-class prediction results.

In this work, we develop, for the first time in the social games domain, a survival ensemble model which provides a comprehensive analysis together with an accurate prediction of churn. For each player, we predict the probability of churning as function of time, which permits to distinguish various levels of loyalty profiles. Additionally, we assess the risk factors that explain the predicted player survival times. Our results show that churn prediction by survival ensembles significantly improves the accuracy and robustness of traditional analyses, like Cox regression.

Index Terms—social games; churn prediction; ensemble methods; survival analysis; online games; user behavior

I. INTRODUCTION

The economics of gaming has changed in the recent years with the widespread adoption of social networks and smartphones, leading to a new type of video games: social games. Social games target a new audience of players: casual gamers, with a new monetization model: free-to-play (F2P or freemium), which now largely dominates all the mobile platforms [2], [14]. The freemium model consists in offering a game for free, and monetizing it by charging for in-game content through in-app purchases.

For social games, player retention is key for a successful monetization, and to increase the social interactions that in turn help to drive the adoption of the game and retain players. In addition, the cost of acquiring new players is ever increasing

Fig. 1. Screenshots of the game chosen to evaluate the churn models, Age of Ishtaria. This game is representative of the successful F2P mobile social role-playing games in Japan. The left panel exhibits characteristic activity in role-playing games, the right panel shows the usual F2P in-app purchases for in-game content.

[14] and can significantly exceed the cost of retaining existing ones.

This study is motivated by the idea that the ability to predict when a player will leave a game allows to take incentive actions to re-engage her and prevent churn, or move her to another game of the company.

Churn prediction has been widely researched in the fields of telecom, finance, retail, pay TV and banking, as shown by the extensive literature review given by [55], [58]. It has also been studied in e-commerce [60], [61] and even in terms of employee retention [51].

In the field of video games, pioneering studies were introduced in [29], [31]. However, they focus on MMORPG (Massively Multiplayer Online Role-Playing Games).

MMORPG have been the first successful type of online social games, however they targeted a narrower audience and they are mainly using a subscription-based monetization

model. This implies the possibility to measure churn as a formal termination of contract, similarly to the sectors mentioned above, at the exception of e-commerce.

Free-To-Play (F2P) monetization, which is the main model used by mobile social games, involves a non-contractual relationship. In this context, churn is not clearly determined by an explicit statement ending a contract. For the most active players, we can define churn as a prolonged period of inactivity. However, the problem slightly differs from the churn in e-commerce. It is indeed always possible for inactive users to come back to an e-commerce website, while inactive mobile players can uninstall a game, which would correspond to a well defined and definitive state of churn. However, this information is normally not available. The definition of churn in non-contractual settings has been discussed in [10]. A comprehensive discussion on the definition of churn for F2P applications is beyond the scope of this paper, and is the subject of dedicated studies [10].

The work presented in [19] is the first study investigating churn prediction in F2P games. [19] introduces a general definition of the problem, a selection of game content independent features and a comparison of classifiers. A second study shown in [48] focuses on the churn prediction of high value players in F2P games. [48] investigates in detail the problem definition and classifier evaluation, though it approaches the problem only from a binary classification point of view. It uses an algorithm that assumes a distribution of data that normally does not fit with the common shape of the churn data. Going further, [49] and [47] try to address the temporality of the data for churn prediction in mobile games.

The work presented in this paper focuses on predicting churn for *high value players* who are commonly called *whales* in the video game industry. A motivation for this focus is that whales behave differently than average players, including in terms of survival curve as we can see in Fig. 2. Since they are often the most active players, i.e. they play nearly every day, we can easily define their churn as a prolonged period of inactivity. Their high level of engagement also allows to collect more data about their activity and makes them more likely to answer positively to actions taken in order to prevent their churn. Finally, from a business perspective, whales, who represent about 0.15% of the players, or 10% of the paying users [28], are particularly important since they are the top spenders who account for 50% of the in-app purchases revenues.

The game chosen for this study, Age of Ishtaria developed by Silicon Studio, is representative of the successful mobile social games and has several million players worldwide.

A. Our contribution

Classical approaches to churn study the problem as a binary classification: whether or not the player connects again to the game (e.g. [3]). Although the binary models are very intuitive, they are not able to predict *when* the player will stop playing

and, moreover, the features are limited to provide static (non-temporal) information.

In order to model the *time* until *churn*, traditional methods like regressions would be appropriate only when all players have stopped playing the game. The challenge arises for data which contains incomplete information about every users, as some of them still play the game.

The present work improves previous studies [19], [48] using an adequate technique that assimilates censored data (observations with incomplete information about churn time) [34] and that captures the temporal dimension of the churn prediction challenge.

Our model based on survival ensembles outputs accurate predictions of *when* players churn, and provides information about the risk factors that affect the exit of players as well. Additionally, the approach suggested in this paper not only gives us a list of possible churners, but also produces, for every player, a *survival* probability function that will let us know how the probability of churning is varying as a function of time. This feature lets us distinguish various levels of loyalty profiles, upcoming, near-future and far-future churners, and the variables that influence this *survival* behavior (considering that a player is *alive* as long as she connects to the social game).

From this *survival* function, the median survival time is extracted and used as a life expectancy threshold. This feature lets us label players as being at risk of churning, take action beforehand to retain valuable players, and ultimately improve game development to enhance player satisfaction.

To the best of our knowledge, we are the first to thoroughly model the prediction of churn by using a survival ensemble approach in the social games sector. Our model improves the accuracy, robustness and flexibility of traditional survival methods, like Cox regression, and has been developed with the goal of being usable in an operational business environment.

II. SURVIVAL ENSEMBLE MODELS

A. Survival Analysis

Survival analysis consists of a set of statistical techniques traditionally used to predict lifetime expectancy of individuals in medical and biological research [15], [26], [35]. This group of methods have also been applied in several industries to predict customer attrition, mainly in telecommunication [36], banking [54] and insurance [16].

Survival analysis focuses on studying the *time* until an *event* of interest happens and its relationship with different factors. Originally in medical research, an event is the failure or death of an individual, however in our case it is the moment when a player leaves the game. The *time-to-event* outcome is also known as *survival time*.

A fundamental characteristic of survival analysis is that data are *censored*. Censoring indicates that observations do not include complete information about the occurrence of the event of interest. It means that for a certain number of players, we do not know the time of event experience (because they did not experience it yet), i.e. measurements only contain information if the event occurs or not before a given time t.

The survival function $S(t)$, which is simply the likelihood that a player will survive at a certain time t, can be estimated through the non-parametric *Kaplan-Meier* estimator [30], where the churn probability can be computed directly from recorded censored survival times.

If k players churn during the period of time T of study at different instants $t_1 < t_2 < t_3 \ldots < t_k$ and, as churn occurrences are supposed to be independent of each other [9], the probabilities of surviving in the game from one time to the next can be multiplied to obtain the *cumulative survival probability*:

$$S(t_j) = S(t_{j-1})\left(1 - \frac{d_j}{n_j}\right) \tag{1}$$

where $S(0) = 1$, with n_j being the number of players alive before t_j, and d_j being the number of events at t_j. We will get as a result a step function that changes its value at the time of each churn.

Further analysis on this topic includes the presence of *competing risks* [43]. They belong to a special class of time-to-event models where there is more than one possible failure event. These alternative events can prevent the observation of the main event of interest. In this study, we focus on the loss of interest in a game, which is the main cause of churn. However, it can happen that a player stops playing the game because she loses her phone, or dies, which are considered as competing risks events.

Additional semi-parametric survival techniques, like the renowned regression method for censored observations, the *Cox proportional-hazards model* [11]–[13], or parametric methods (e.g. accelerated failure time models [38]), are valuable tools to investigate the impact of multiple covariates. The covariates or predictors are expected to be correlated with the player's reason for quitting the game.

Following Cox proportional-hazards model, the estimated hazard for k individual players and p covariate vectors x_k takes the form

$$h_k(t) = h_0(t)\exp\left(\beta_1 x_{k,1} + \ldots + \beta_p x_{k,p}\right), \tag{2}$$

where the hazard function $h_k(t)$ is dependent on the baseline hazard $h_0(t)$ and the features $\beta_p x_{k,p}$. The Cox regression is not assumed to follow a particular statistical distribution. It is fitted based on the data and it solves the censoring problem by maximizing the partial likelihood.

The Cox model and its extensions [56] allow regressions to work with censored data, and they permit an intuitive interpretation of the impact of the features. However, these techniques assume a fixed link between the output and the variables (assuming them additive and constant over time). This requires an explicit specification of the relationship by the researcher, and involves important efforts in terms of model selection and evaluation. In spite of their semi-parametric nature, these models present difficulties to scale with big data problems, and alternative *regularized* versions of Cox regression [39] have been proposed to amend this. Nevertheless, they are still based on restrictive assumptions that are not easy to fulfill.

In the parametric approaches, like the accelerated failure time models [38], the type of the distribution is previously determined (e.g. Weibull, lognormal, exponential). Though, these methods are suboptimal because it is uncommon that the data follow these specific distribution shapes.

In the present paper, we address the drawbacks mentioned above by applying machine learning algorithms to censored data problems.

B. Survival Trees and Ensembles

1) Decision Trees: Originally presented in [41], decision trees became popular in the 1980s, when the most relevant algorithms for *Classification and Regression Trees* (CART) were introduced by [7], [44], [50].
Classification and regression trees are non-parametric techniques where the basic idea is to split the feature space recursively, to group subjects with homogeneous characteristics and to separate those with bigger differences based on the outcome of concern. In order to perform the nodes classification and maximize homogeneity within the nodes, a measure called *impurity* must be minimized. Common examples of impurity measure are cross-entropy or sum of squared errors. For example, considering a binary split and given a continuous variable X, the split can be performed if $X \leq d$ is fulfilled, with d being a constant.

2) Survival trees: Survival trees are constructed as a set of binary trees that grow by recursive partitioning of the sample space χ, where the q_i tree nodes are subspaces of χ. The tree splitting starts in the root node, which concentrates all the data. Based on a survival statistical criterion, such as the cumulative hazard function or Kaplan-Meier estimates, the root node is then divided into two daughter nodes. The principle for partitioning these two branches is to maximize the survival difference between two groups of individuals, which are compressed in the two daughter nodes, maximizing the homogeneity among nodes, based on survival experience.

The first idea of using tree-based methods for censored data was initially introduced in [8] and [37]. The first *survival tree* as we know was presented in [17], where a Kaplan-Meier estimator survival function was computed at every node as a discrepancy measure using Wasserstein metrics. For a comprehensive review about different types of survival trees, check [4].

The best split is achieved by exploring all combinations, considering all the x_i predictor variables and all the possible splits, in order to maximize the survival difference. This way, subjects with similar survival characteristics are grouped together. As long as the tree grows, the difference between branches increases, and individuals are gathered in nodes with more homogeneous groups in terms of survival behavior.

Despite being a powerful classification tool which is able to model censored data, employing a single tree can produce instability in its predictions. It means that if small changes in data arise, the prediction can differ among computations (the divergences are mainly related with the prediction of risk

factors) [33]. This drawback will be fixed if we execute an ensemble of them, instead of using one single tree.

3) Survival ensembles: Using an ensemble of models, instead of a single one, is an accurate prediction tool firstly suggested by [5], [6] with the well-known *random forest*. Ensembles of tree-based models achieve outstanding predictions in real-world applications [62].

Survival forests are ensemble-based learning methods where the underlying algorithm is a kind of survival tree. A survival ensemble lies in growing a set of survival trees, instead of a single one. The two main survival ensemble techniques are *random survival forest*, presented in [27], and *conditional inference survival ensembles*, developed by [24], based on their previous work introduced in [21], [25].

The conditional inference survival ensembles is the method chosen for the predictions shown in Section III. The conditional inference survival ensemble technique uses a weighted Kaplan-Meier function based on the measurements used for the training. The ensemble survival function [40] can be summarized by

$$S^{conditional}(t|x_i) = \prod \left(1 - \frac{\sum_{n=1}^{N} T_n(dt, x_i)}{\sum_{n=1}^{N} Q_n(t, x_i)} \right) \quad (3)$$

where n indicates the number of trees within the ensembles, with $n = 1, \cdots, N$, and x_i being the covariates. Therefore, in the node where x_i is located, T_n accounts for the uncensored events until time t, and Q_n counts the number of individuals at risk at time t. Moreover, conditional inference survival ensembles introduces additional weight to the nodes where there are more subjects at risk. It uses linear rank statistics as splitting criterion to grow the trees.

In contrast, random survival forests [27] are based on Nelson-Aalen estimates (instead of using Kaplan-Meier estimates). The maximum of the log-rank statistical test is used in every node as split criterion, which leads to biased results in favor of covariates with many splits.

Conditional inference survival ensembles is a promising approach to deal with the censoring nature of churn prediction. It is a flexible method compared to the traditional statistical Cox regression model and it solves the instability that is present in survival trees. In the selected method for the churn study, overfit is not present in its estimates and provides robust information about the variable importance. This fixes the random survival forest problem [59] of being biased towards predictors with many splits or missing data.

III. DATASET

We collected data from a major mobile social game between October 2014 and February 2016. Several churn predictors or risk factors were investigated.

We investigated mainly game-independent features, i.e. features that are not related to the game mechanics and can be measured in any game. This allows us to build a game-independent churn prediction model that can be applied to other games.

Additionally, we want to implement our model in a data science product running in an operational business environment. Thus, the feature selection takes into account limitations in terms of memory and processing capabilities that might not be considered in a pure research environment.

- Player attention: the time component of the player accessing the game.
  - Time spent per day in the game, including averages over the first weeks and moving average over the last weeks.
  - Lifetime: number of days since registration until churn, in case the player churns.
- Player loyalty: the frequency of the player access to the game.
  - Number of days with at least one playing session.
  - Loyalty index: ratio of number of days played, divided by lifetime.
  - Days from registration to first purchase.
  - Days since last purchase.
- Playing intensity: the quality of the playing sessions, i.e. how a player interacts with the game.
  - Number of actions.
  - Number of sessions.
  - Number and amount of in-app purchases.
  - Action activity distance: Euclidean distance between the average number of actions over the lifetime and the average number of actions over the last days.
- Player level: the value of this variable and its evolution depends on the game. However, the concept of level is present and measurable in the majority of games, and can be then considered as a game-independent predictor that can be used in our model and applied to most other mobile social games.

We investigated some game-dependent features, that we ultimately did not keep in our model, such as:

- Participation in a *guild*. Guild is a social feature, sometimes called *union* or *clan*, specific to some social games, allowing to play in collaboration with other players. This predictor turned to be inapplicable to our problem as the whales, who are the focus of this study, have an homogeneous behavior in terms of participation to the social features of the game.
- Measure of number of actions by category (shop, battle, mission, ...). This is specific to the game studied. Though, it does not bring more relevant information than the higher-level and game-independent measure of the total number of actions.

IV. MODELING

A. Churn definition

As it was explained in Section I, the definition of churn in F2P games is not straightforward. In this study, we consider that a player has churned if she does not connect to the game for 10 consecutive days. Our measurements confirm that the
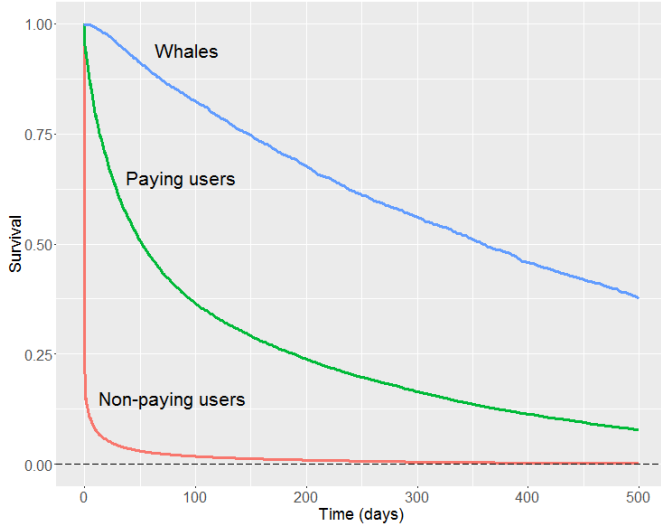
Fig. 2. Cumulative survival probability (Kaplan-Meier estimates) as a function of time, in days. The whole dataset is considered, and the results are stratified by whales (high-value players), paying users, and non-paying users.

whales who went through a period of 10 days of inactivity become mostly inactive: they either permanently exit the game or their activity becomes neglectable. Indeed the purchase activity of whales after coming back to the game following a period of 10 days of inactivity represents only 1.4% of the revenues generated by this category of players.

Traditional churn analysis focuses on predicting whether or not a user is going to exit the game, i.e. the response is a binary variable: yes or no. However, with this approach, we do not know when a player is going to stop connecting to the game. Conventional churn prediction is solved from a static point of view, a binary classification problem.

Our work focuses on *when* churn will happen. We model the churn behavior from the perspective of survival analysis, and we treat the prediction of churners as a censored data problem where the outcome of our model is the continuous time - the *time-to-exit* the game. We have used the algorithm of survival ensembles within the conditional inference framework from [24], presented in Section II-B3. This study uses a learning sample of $n = 2500$ whales.

B. Kaplan-Meier estimates

We visualize the churn problem by plotting Kaplan-Meier (K-M) survival curves stratified by whales, normal paying users, and non-paying users. In order to perform the K-M survival analysis, we take a sample of 1.500.000 players.

Fig. 2 provides a graphical representation of the K-M survival curves for different kind of players based on their paying behavior, distinguishing among whales (high-value players), paying users, and non-paying users. Fig. 2 shows different survival patterns for each group. The estimated survival for non-paying users is much lower than the one of paying users (including both whales and non-whales paying users). Approximately 80% of the non-playing players have churned

the first day they connected to the game. This contrasts with the 20% churn rate of the whales after 100 days.

C. Churn model as a censored data problem

In the present work, the authors propose conditional inference survival ensembles [24] to model game churn.

Survival ensembles with 1000 conditional inference trees are used as a base learner to predict the exit time of whales from the game. Fig. 3 shows how conditional inference trees work. It illustrates a simple partition with two terminal nodes. In each terminal node, a Kaplan-Meier survival curve represents the group of players included in the node classification. In this example we can observe the differences between the survival profiles that characterize every node. In Fig. 3, the root node variable is the last level the player reached in the game. Two daughter nodes partitions grow from it: one also based on the level and another based on the number of days since the player did the last in-app purchase *daysLastPurchase*.

The overall survival time is the outcome of this model. Fig. 4 summarizes the most significant predictors included in the survival ensemble model for right-censored observations. The variable importance is computed using the integrated Brier score (IBS) [18], and the feature selection is performed based on it. Other survival ensemble methods, like [27], are not as robust as the technique employed in this work [24] in terms of variable selection and therefore in terms of computation of variable importance. The variable importance is normally biased in favor of the predictors with many splits. Conditional inference survival ensembles are constructed based on unbiased trees, avoiding this problem [23].

The resulting prediction of this model contains, for each player, a survival function indicating the probability of churn as a function of time since the registration in the game. Fig. 5 illustrates a sample of four Kaplan-Meier survival functions for four new players. In Fig. 5, we can observe the probability of churn for every single player (y-axis) as a function of time in days (x-axis). In this example, we distinguish different player profiles and survival behaviors:

- The two first plots starting from the left show the survival probability curves of two players who are going to churn soon.
- The third plot starting from the left shows the survival probability curve of a player who is expected to churn but not in the near future.
- The last plot starting from the left shows the survival probability curve of a very loyal player.

For every player, a different survival function will be computed as a result of our model.

Fig. 5 highlights the capability of our model to classify and predict loyalty for every player, taking into account the temporal dimension. Additionally, the median survival time, which is the time when the percentage of surviving in the game is 50%, is used as a time threshold to categorize a player as being at risk of churning.
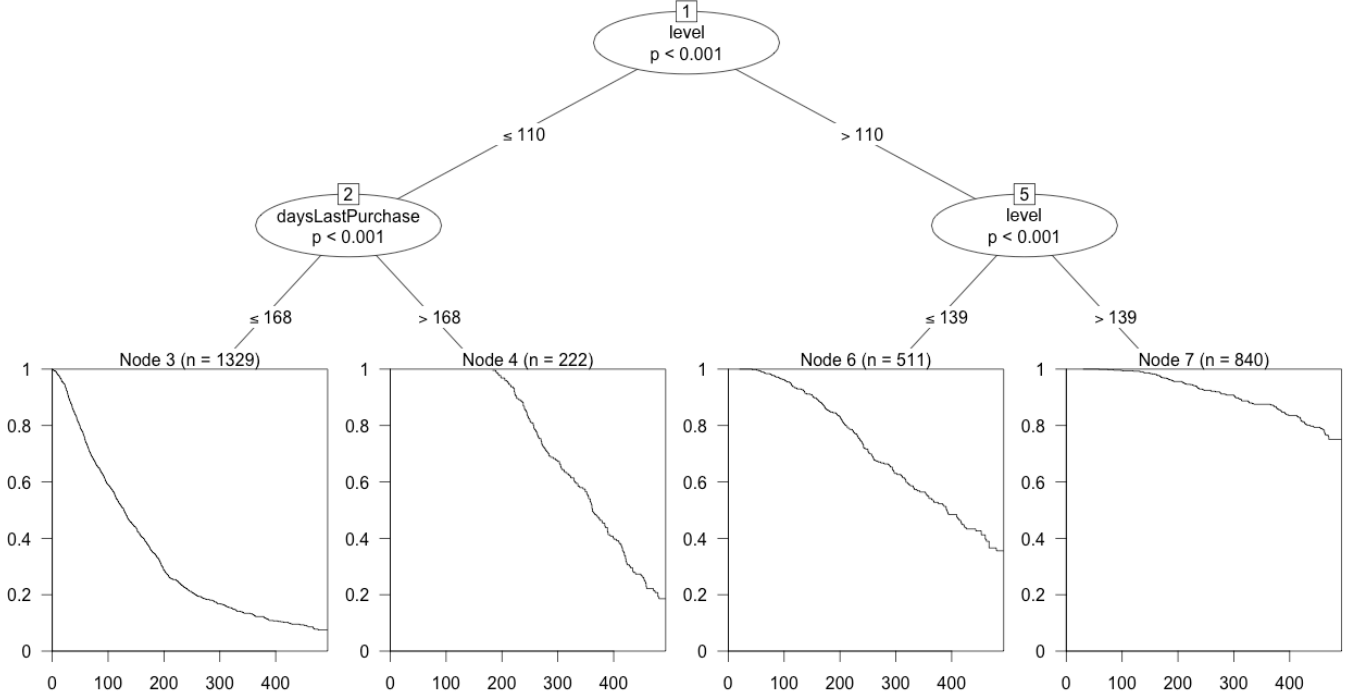
Fig. 3. Conditional inference survival tree partition with the estimated Kaplan-Meier survival curve. The K-M estimates represent the survival probability (in days), which characterizes the players placed in every terminal node.
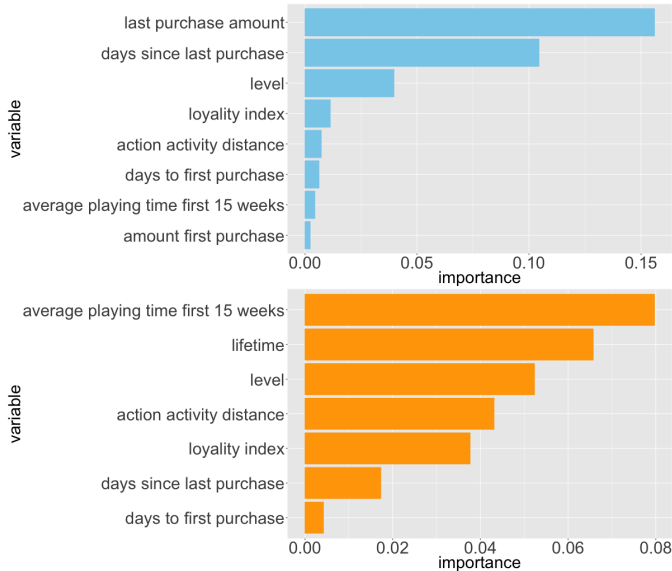

Fig. 4. Variable importance (relative ranking of significant factors) to predict time-to-churn with conditional inference survival ensemble model in the upper panel, and for the binary output in the lower panel.

D. Model validation

Because of the nature of censoring, the standard methods of visualizing and evaluating prediction performances are not suitable [40]. Fig. 7 shows the fit of the proposed conditional inference survival ensemble method and the selected Cox regression (using the same predictors). The conditional inference survival ensemble model exhibits a reasonable agreement between measured and predicted survival times, both in the scatterplot and the mean-difference plot evaluation. We can observe in the lower plots that Cox regression performs worse than the ensemble model in terms of predictive ability. As it can be observed in Fig. 7, there is a higher concentration of data at the beginning of the study. This is due to the fact that we work with censored data and do not follow a normal distribution. Hence, the longer the time of study grows, the less information we have, as there are many whales who have not experienced the *event* yet because they are still connecting to the game. This evidence is reflected in the cumulative survival distribution for whales shown in Fig. 2. Thus, as long as the censoring rate grows, the prediction capability diminishes.

Fig. 6 depicts the cumulative prediction error curve for the survival ensemble and the Cox regression model. The integrated Brier score is an evaluation measure developed for survival analysis [18], [40]. We use it to establish the summary of the error estimation for the two survival-time analysis outputs. The error evaluation has been performed based on bootstrap cross-validation with replacement. This technique estimates the prediction error splitting the measurements in many bootstrap training and test samples. Then, the models are trained and tested with multiple sets of bootstrap samples. Fig. 6 exhibits the bootstrap cross-validated prediction error curves for 1000 samples.

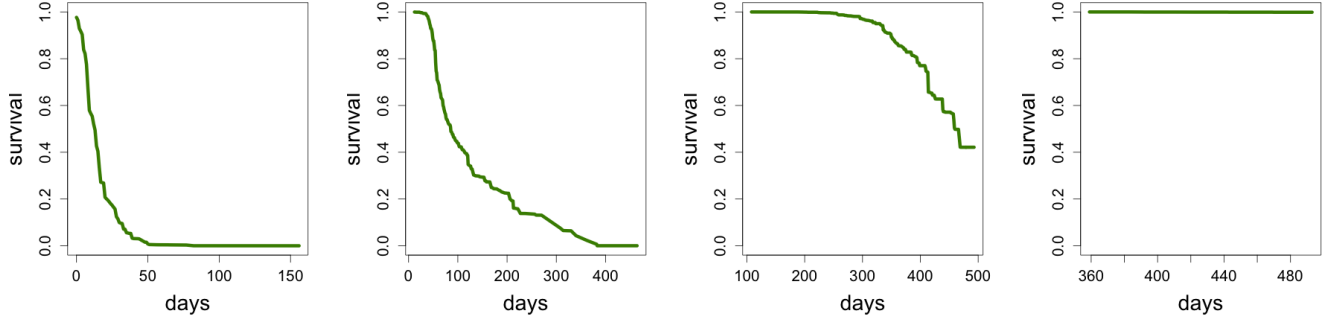Fig. 6 basically supports what Fig. 7 shows, as the

Fig. 5. Predicted Kaplan-Meier survival curves as a function of time (days) for new players. Predictions performed with conditional inference models for survival ensembles.
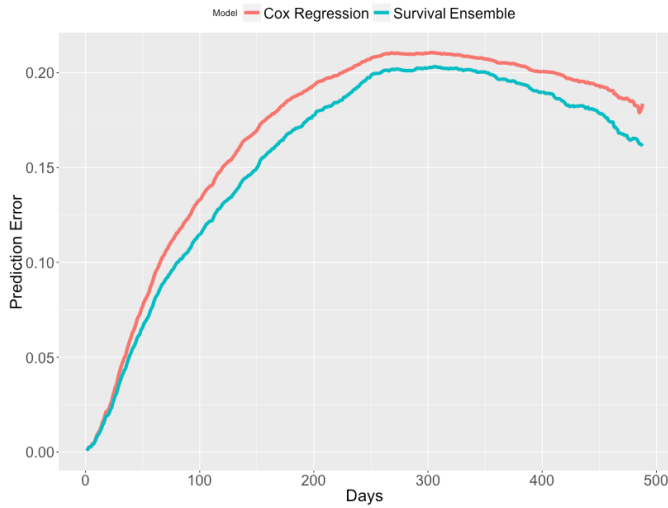

Fig. 6. 1000 bootstrap cross-validation error curves for the survival ensemble model and Cox regression.

TABLE I
PREDICTION ERROR SCORE FOR THE SURVIVAL MODELS WITH
INTEGRATED BRIER SCORE (IBS)

Model	IBS
Survival Ensemble	0.158
Cox regression	0.169
Kaplan-Meier	0.199

ensemble-based approach improves accuracy over the Cox model, cf. Table I. The prediction error function reaches the maximum at the median survival time of 304 days and 306 days, for the Cox regression (error value of 0.21) and ensemble model (error value of 0.20), respectively.

Additional validation tests have been performed to compare the accuracy of the two models. A paired t-test (Welch Two Sample t-test) [42] is used to estimate whether the prediction ability of a model is statistically significant from another. The t-test has been performed using a confidence interval of 95%. According to the t-test, survival ensemble model is statistically significant, i.e. p-value ≤ 0.05 . We obtain the following values: $t = 3.56$ and $p = 0.00039$.

V. COMPARISON WITH OTHER MODEL APPROACHES

We include in this section a binary classification model of churners. Although we think that modeling churn as a censoring problem is the adequate approach, the binary prediction perspective also brings us interesting information. The binary response model provides useful insight for a very short-term prediction. It is easy to interpret and to implement.

Although we use the same algorithm of conditional inference ensembles, the outcome differs. A binary variable denoting if a player churns or not is the response of the classification model, i.e. yes or no. We trained the binary model with several sets of features to obtain the final list of attributes shown in Fig. 4. We highlight the contrasting results obtained during the evaluation of the variable impact between the survival model and the binary classification. It reflects the nature of different ways of modeling and therefore of the prediction results.

A comparison study with other binary classification methods was performed in order to support the results obtained with the binary approach of the churn analysis. For this study, we select several algorithms as binary classificators: SVM, naive bayesian classifier and a decision tree. A detailed and complete explanation about the techniques used here can be found in [20].

The fit of the ensemble is summarized in Table II, where we compare our results with other classification methods. It indicates a good agreement between observed and predicted churners with an AUC (area under ROC curve) of 0.96. Although the other techniques also perform very accurately, they possess some drawbacks. SVM also have a high score in terms of AUC, but they are considered as *black boxes* because it requires significant effort to extract the relationship between the input variables and the output [32].

The techniques applied above are powerful tools to solve regression or classification problems. However, in their original, form they cannot handle the assimilation of information from censored data. Hence, in order to apply these methods
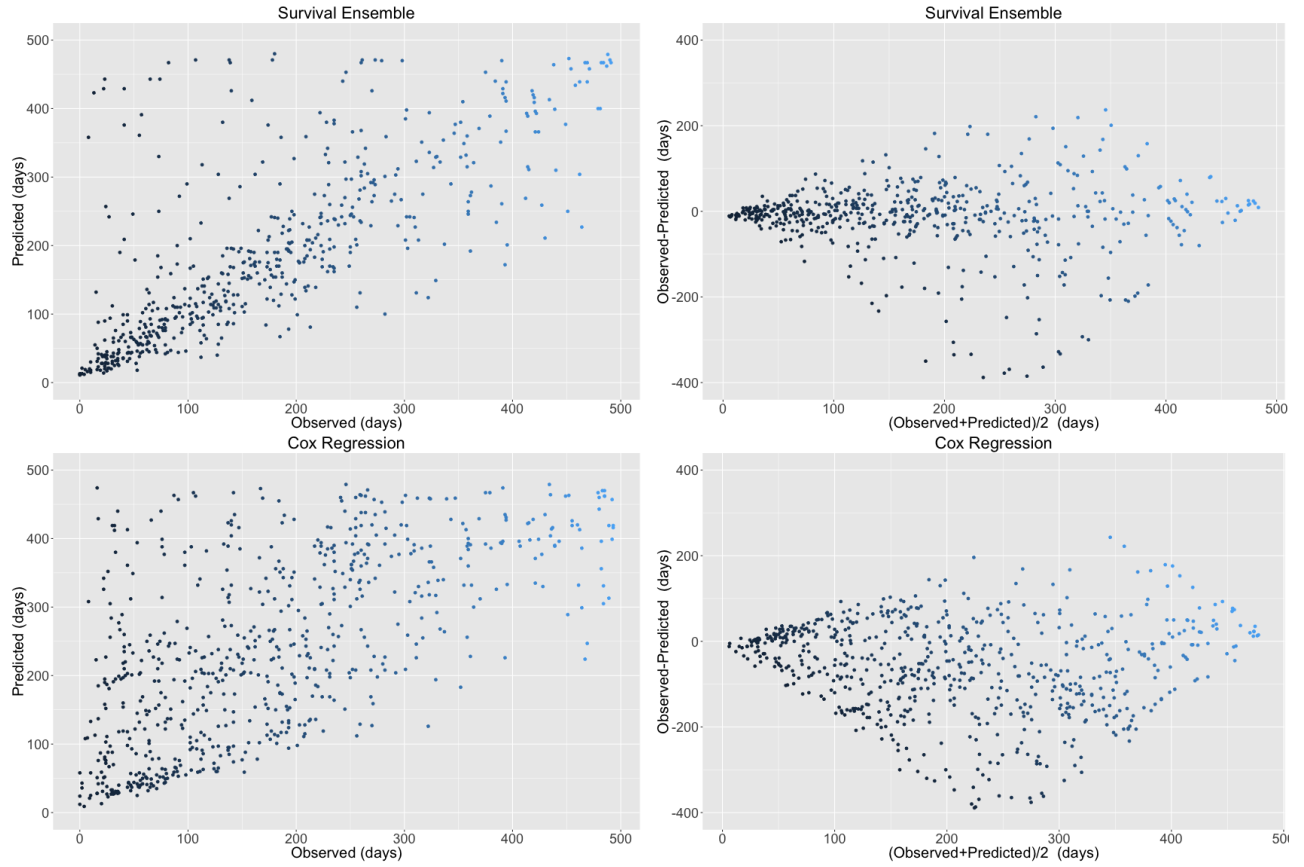
Fig. 7. Scatterplot (left) and mean-difference plot (right) of observations and predictions of median survival times. The dark blue dots correspond to shorter lifetimes (in days) of players, soft blue dots reflect players with longer lifetimes. Upper panel evaluates the survival ensemble results, and lower panel compares the Cox regression analysis.

Model	AUC
Survival Ensemble	0.960
Support Vector Machines	0.940
Naive Bayesian	0.900
Decision Tree	0.934

for survival analysis responses, an adequate modification of the algorithm or a proper transformation of the data must be performed beforehand.

VI. SUMMARY AND CONCLUSION

The focus of this research is to find an appropriate technique to model player churn, which has been an open problem within the community. Furthermore, this work presents steps towards the challenging goal of understanding the most valuable players in social games.

The authors propose the application of a state-of-the-art algorithm: conditional inference survival ensembles [24], to predict the time-to-churn and the survival probability of players in social games in terms of game lifetime.

We look for a method that is able to make predictions in an operational business environment and that easily adapts to different kinds of games, players, and therefore distribution of the data. This is the main motivation: we need a flexible technique that does not require a previous manipulation of the data and that is able to deal efficiently with the temporal dimension of the churn prediction problem. Conditional inference survival ensembles were evaluated to this purpose and compared with traditional survival methods like Cox regression.

Conditional inference survival ensembles provided more accurate and more stable prediction results than traditional approaches. The proposed method is unbiased, does not overfit [24], and provides us with robust information about the risk factors that influence players to abandon the game.

The predictions we have obtained provide the business users and game developers with useful and easy-to-interpret player information. The results directly impact the game business, improving the knowledge about whales behavior, discovering new playing patterns as a function of time, and classifying social gamers by risk factors of churn.

Further on-going work in this direction is the improvement of the accuracy in the prediction of the time-to-churn for players who stay longer in the game. To achieve this, we will continue researching significant features to discover new

playing patterns. A promising direction would be to study predictors based on more complex measures of the social activity than the one used in this study.

VII. Software

All the analysis were performed with R version 3.1.2 for Linux, using the following packages from CRAN: *party* version 1.0 [22], [23], *survival* version 2.38 [57], *ROCR* 1.0 [52], [53], *Core Learn* version 1.47 [45], [46] and *cvTools* version 0.3.2 [1].

Acknowledgements

References

[1] A. Alfons. cvTools: Cross-validation tools for regression models. *R package version 0.3*, 2(5), 2012.

[2] A. Annie. App annie and IDC mobile app advertising and monetization trends, 2013.

[3] J. Banasik, J. N. Crook, and L. C. Thomas. Not if but when will borrowers default. *Journal of the Operational Research Society*, pages 1185–1190, 1999.

[4] I. Bou-Hamad, D. Larocque, H. Ben-Ameur, et al. A review of survival trees. *Statistics Surveys*, 5:44–71, 2011.

[5] L. Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

[6] L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[7] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. *Classification and regression trees*. CRC press, 1984.

[8] A. Ciampi, R. Bush, M. Gospodarowicz, and J. Till. An approach to classifying prognostic factors related to survival experience for non-hodgkin's lymphoma patients: Based on a series of 982 patients: 1967–1975. *Cancer*, 47(3):621–627, 1981.

[9] L. S. A. D. Clark TG, Bradburn MJ. Survival analysis part i: Basic concepts and first analyses. *British Journal of Cancer*, 89(2):232–238, 2003.

[10] M. Clemente-Císcar, S. San Matías, and V. Giner-Bosch. A methodology based on profitability criteria for defining the partial defection of customers in non-contractual settings. *European Journal of Operational Research*, 239(1):276–285, 2014.

[11] D. R. Cox. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(2):187–220, 1972.

[12] D. R. Cox and D. Oakes. *Analysis of survival data*, volume 21. CRC Press, 1984.

[13] C. R. David. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, 34:187–220, 1972.

[14] T. Fields. *Mobile and Social Game Design: Monetization Methods and Mechanics*. CRC Press, 2 edition, 2014.

[15] T. R. Fleming and D. Lin. Survival analysis in clinical trials: past developments and future directions. *Biometrics*, 56(4):971–983, 2000.

[16] L. Fu and H. Wang. Estimating insurance attrition using survival analysis. *Table of*, page 55.

[17] L. Gordon and R. Olshen. Tree-structured survival analysis. *Cancer treatment reports*, 69(10):1065–1069, 1985.

[18] E. Graf, C. Schmoor, W. Sauerbrei, and M. Schumacher. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine*, 18(17-18):2529–2545, 1999.

[19] F. Hadiji, R. Sifa, A. Drachen, C. Thurau, K. Kersting, and C. Bauckhage. Predicting player churn in the wild. In *Computational Intelligence and Games (CIG), 2014 IEEE Conference on*, pages 1–8. IEEE, 2014.

[20] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference and prediction*. Springer, 2 edition, 2009.

[21] T. Hothorn, P. Bühlmann, S. Dudoit, A. Molinaro, and M. J. Van Der Laan. Survival ensembles. *Biostatistics*, 7(3):355–373, 2006.

[22] T. Hothorn, K. Hornik, C. Strobl, and A. Zeileis. Party: A laboratory for recursive partytioning, 2010.

[23] T. Hothorn, K. Hornik, C. Strobl, A. Zeileis, and M. T. Hothorn. Package 'party'. *Package Reference Manual for Party Version 0.9-998*, 16:37, 2015.

[24] T. Hothorn, K. Hornik, and A. Zeileis. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3):651–674, 2006.

[25] T. Hothorn, B. Lausen, A. Benner, and M. Radespiel-Tröger. Bagging survival trees. *Statistics in medicine*, 23(1):77–91, 2004.

[26] P. Hougaard. Fundamentals of survival data. *Biometrics*, 55(1):13–22, 1999.

[27] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, and M. S. Lauer. Random survival forests. *The annals of applied statistics*, pages 841–860, 2008.

[28] E. Johnson. A long tail of whales: Half of mobile games money comes from 0.15 percent of players, 2014. *http://recode.net/2014/02/26/a-long-tail-of-whales-half-of-mobile-games-money-comes-from-0-15-percent-of-players*.

[29] D. G. Jun Ding and X. Chen. Alone in the game: Dynamic spread of churn behavior in a large social network a longitudinal study in MMORPG. *falta*, 24(2):123–140, 1996.

[30] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

[31] J. Kawale, A. Pal, and J. Srivastava. Churn prediction in MMORPGs: A social influence based approach. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, volume 4, pages 423–428. IEEE, 2009.

[32] M. Kretowska. The influence of censoring for the performance of survival tree ensemble. *Springer Berlin Heidelberg*, pages 524–531, 2010.

[33] M. Kretowska. *Artificial Intelligence and Soft Computing: 13th International Conference, ICAISC 2014, Zakopane, Poland, June 1-5, 2014, Proceedings, Part I*, chapter Comparison of Tree-Based Ensembles in Application to Censored Data, pages 551–560. Springer International Publishing, Cham, 2014.

[34] S. Lagakos. General right censoring and its impact on the analysis of survival data. *Biometrics*, pages 139–156, 1979.

[35] J. Li and S. Ma. *Survival analysis in medicine and genetics*. CRC Press, 2013.

[36] J. Lu. Predicting customer churn in the telecommunications industry - An application of survival analysis modeling using SAS. *SAS User Group International (SUGI27) Online Proceedings*, pages 114–27, 2002.

[37] E. Marubini, A. Morabito, and M. Valsecchi. Prognostic factors and risk groups: some results given by using an algorithm suitable for censored survival data. *Statistics in medicine*, 2(2):295–303, 1983.

[38] E. Marubini and M. G. Valsecchi. *Analysing survival data from clinical trials and observational studies*. Wiley-Interscience, 2004.

[39] S. Mittal, D. Madigan, R. S. Burd, and M. A. Suchard. High-dimensional, massive sample-size cox proportional hazards regression for survival analysis. *Biostatistics*, page kxt043, 2013.

[40] U. B. Mogensen, H. Ishwaran, and T. A. Gerds. Evaluating random forests for survival analysis using prediction error curves. *Journal of statistical software*, 50(11):1, 2012.

[41] J. N. Morgan and J. A. Sonquist. Problems in the analysis of survey data, and a proposal. *Journal of the American statistical association*, 58(302):415–434, 1963.

[42] N. I. of Standards, T. (US), C. Croarkin, P. Tobias, and C. Zey. *Engineering statistics handbook*. The Institute, 2001.

[43] R. L. Prentice, J. D. Kalbfleisch, and A. V. Peterson. The analysis of failure times in the presence of competing risks. *Biometrics*, 34:541–544, 1978.

[44] J. R. Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.

[45] M. Robnik-Sikonja and P. Savicky. CORElearn - classification, regression, feature evaluation and ordinal evaluation. *The R Project for Statistical Computing*, 2012.

[46] M. Robnik-Sikonja, P. Savicky, and M. M. Robnik-Sikonja. Package 'CORElearn', 2013.

[47] P. Rothenbuehler, J. Runge, F. Garcin, and B. Faltings. Hidden markov models for churn prediction. In *SAI Intelligent Systems Conference (IntelliSys), 2015*, pages 723–730. IEEE, 2015.

[48] J. Runge, P. Gao, F. Garcin, and B. Faltings. Churn prediction for high-value players in casual social games. In *Computational Intelligence and Games (CIG), 2014 IEEE Conference on*, pages 1–8. IEEE, 2014.

[49] A. Saas, A. Guitart, and A. Perianez. Discovering playing patterns: Time series clustering of free-to-play game data. *Computational Intelligence and Games (CIG), 2016 IEEE Conference on*, 2016.

[50] S. L. Salzberg. by j. ross quinlan. morgan kaufmann publishers, inc., 1993. *Machine Learning*, 1:6, 1994.

[51] V. Saradhi and G. K. Palshikar. Employee churn prediction. *Expert Systems with Applications*, 38(3):1999–2006, 2011.

[52] T. Sing, O. Sander, N. Beerenwinkel, and T. Lengauer. Package 'ROCR': visualizing the performance of scoring classifiers, 2007. *http://rocr.bioinf.mpi-sb.mpg.de*.

[53] T. Sing, O. Sander, N. Beerenwinkel, and T. Lengauer. ROCR: visualizing classifier performance in R. *Bioinformatics*, 21(20):3940–3941, 2005.

[54] M. Stepanova and L. C. Thomas. Survival analysis methods for personal loan data. *Operations Research*, 50(2):277–289, 2002.

[55] A. Tamaddoni Jahromi, M. M. Sepehri, B. Teimourpour, and S. Choobdar. Modeling customer churn in a non-contractual setting: the case of telecommunications service providers. *Journal of Strategic Marketing*, 18(7):587–598, 2010.

[56] Terry M. Therneau and Patricia M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. Springer, New York, 2000.

[57] T. M. Therneau and T. Lumley. Package 'survival', 2015.

[58] W. Verbeke, D. Martens, C. Mues, and B. Baesens. Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications*, 38(3):2354–2364, 2011.

[59] M. N. Wright, T. Dankowski, and A. Ziegler. Random forests for survival analysis using maximally selected rank statistics. *arXiv preprint arXiv:1605.03391*, 2016.

[60] S. Yoon, J. Koehler, and A. Ghobarah. Prediction of advertiser churn for google adwords. In *JSM proceedings*, 2010.

[61] X. Yu, S. Guo, J. Guo, and X. Huang. An extended support vector machine forecasting framework for customer churn in e-commerce. *Expert Systems with Applications*, 38(3):1425–1430, 2011.

[62] C. Zhang and Y. Ma. *Ensemble machine learning*. Springer, 2012.