

Retention and Progression: Seven Months in World of Warcraft

Thomas Debeauvais
Cristina V. Lopes
UC Irvine
Irvine, CA
{tdebeauv, lopes}@uci.edu

Nick Yee
Nicolas Ducheneaut
Ubisoft*
contact@nickyee.com
ducheneaut@gmail.com

ABSTRACT

Knowing which players will stop playing is important for on-line game companies. Players generate massive amounts of behavior data every day. Can we mine this data to predict if they will churn? Is churn related to the pace of progression through game content? We collected demographic and motivational data from a survey of 1350 players from China and North-America, and matched it with raiding and player-versus-player data collected from the game between December 2011 and June 2012.

We find that the ratio of active player base who raids remains constant around 50% across all seven months. However, the active player base of June raids half as much as its December counterpart. Our results also indicate that Chinese players are more focused, while North-Americans more adverse to difficulty. While 10% of the player base churn every month, 5% come back, thus netting a 5% active player loss per month. A simple regression model predicting player churn the following month using only three in-game features achieves 0.90 recall but only 0.28 precision, suggesting that churn remains challenging to predict.

Categories and Subject Descriptors

K.8.0 [Personal Computing]: General – Games

Keywords

MMOs; game design; retention; churn; progression;

1. INTRODUCTION

Long-running games such as Massively Multiplayer Online games (MMOs) have the particularity of receiving new content regularly. Players can become bored once they have explored all the game's content, and may stop playing. Therefore game companies want to know how fast players consume

*Nick Yee and Nicolas Ducheneaut were working at the Palo Alto Research Center (PARC) at the time of the study.

content so they can schedule the release of new content accordingly. Once new content goes live, the overall player base may progress through it faster or slower than the developers expected. Knowing whether the player base responds well to new content is so important that when the developers realize there is a problem, they often release a “hotfix” patch addressing it the same day.

Beyond content consumption, MMO developers are also interested in two other player behaviors: presence and churn. Presence is the number of people who play at a given time. Churn is the number of players who leave the game. Developers want to quantify these two behaviors for two reasons. First, presence and churn drive virality and adoption. For example, an active player may talk about the game to her friends, or contribute to the positive social atmosphere in the game. If she exerts a lot of influence in the game and churns, she may pull with her other players, and revenues may decrease significantly. Second, predicting presence can help keep the in-game economy in check. For example, if too many people play, and the sinks do not match the faucets, then the in-game economy quickly collapses. If too few people play, and the main way for players to obtain equipment is through an auction house, then there may be a shortage of equipment.

Tracking presence, churn, and content consumption has become so important that data-mining has bloomed in the game industry and academic game studies. In the industry, Valve has been conducting extensive physiological player behavior research [1]. Microsoft Games Studios collected and aggregated achievement data to compare player progression across many Xbox 360 games [19]. And some third-party consulting companies such as Ninja Metrics are now analyzing the in-game social networks to predict churn. In academia, several studies have focused on progression, presence, or churn. One study on progression found that guilds can raid dungeons in World of Warcraft at different speeds [3]. Another study found that children are less likely to explore optional content than adults in educational game [17]. Another looked at death and level completion time in Tomb Raider: Underworld [12]. Focusing on churn, a study conducted an exploratory analysis of demographic features related to taking a break in WoW [11]. Two other studies in EverQuest II focused on separating churning from non-churners using decision trees [7] and predicting churn from the game's social network [14].

In this paper, we ask: Is there a relationship between content consumption and presence? Which demographic factors influence content consumption and presence? Can we predict presence from one month to the next? We look at an MMO called World of Warcraft (WoW). WoW is a great example because it has set the standards in the industry [22], it is available in most regions of the world, it released new content in November 2011, and yet its subscriber base dropped from 10.3 million subscribers in November 2011 [15] to 9.1 million in June 2012 [16]. Given that each subscriber pays around \$10 per month to play, the game’s profits decreased significantly in only seven months. We first introduce the main features of the game and our study protocol, then look at presence across various demographics, and progression in dungeon raids. We finally compare two models predicting churn, and conclude with recommendations for online games.

2. METHODS

In this section, we first introduce World of Warcraft (WoW) and the new content introduced by patch 4.3, which is the focus of this paper. Then we review existing MMO player typologies, and pick one that is appropriate for studying WoW players. And then we detail the protocol used in recruiting participants and the data collected from them.

2.1 World of Warcraft 4.3

WoW is a medieval fantasy MMO developed and released by Blizzard Entertainment. Blizzard launched WoW in November 2004. Players can wander the world (made of several continents each cut in dozens of zones), kill monsters, loot items, engage in player-versus-player fights (called PvP), trade items, bid or place auctions for items, and socialize. Players can kill monsters by themselves in the open world, or as part of a raid of 5, 10, or 25 players trying to defeat particularly difficult boss monsters in a dungeon. Players can participate in several PvP activities: duels; arenas of 2v2, 3v3, or 5v5; battlegrounds of 10v10 or 40v40; and zone-wide fighting against players of the opposite faction. In WoW, Blizzard has emphasized accessibility (easy to learn) and depth (difficult to master) [18].

Blizzard released expansion 4.0 titled “Cataclysm” in December 2010. Among other things, this expansion raised the maximum level of a character to 85. Blizzard launched a series of three content patches for Cataclysm, the last of which is patch 4.3 “Hour of Twilight”. Patch 4.3 launched on November 29, 2011, and remained the latest content addition in WoW until expansion 5.0 in September 2012.

Patch 4.3 started PvP season 10, which reset the PvP ratings of all characters to zero. As of raiding, The Firelands (FL) was the hardest dungeon raid in the game before patch 4.3 launched. FL provided the best character equipment and contained seven boss monsters: Alysrazor and three others are outdoors and can be approached in any order, then Baleroc, Staghelm, and Ragnaros must be defeated in that order. FL was restricted to characters who reached the level cap of 85. In fact, a lot of the game content is reserved for characters who have reached the highest level. This part of the game in which progression is no longer the goal is called endgame. Patch 4.3 introduced a new endgame dungeon raid called Dragon Soul (DS) also restricted to characters

level 85, and providing even better equipment than FL. DS features eight sequential boss encounters, with Morchok the first and easiest, Ultraxion the fifth and of medium difficulty, and Madness of Deathwing the last and most difficult. In both DS and FL, players can plan their raid with fellow guild members in normal or heroic difficulty. In heroic difficulty, monsters have more health points, inflict more damage, and are overall more difficult. To try a dungeon in heroic difficulty, players must first complete it in normal difficulty. In DS, they may also raid on-demand with random players in the “Looking for Group” (LFG) difficulty. Both DS and FL are available for 10- or 25-player groups. Once players have defeated a boss, they can skip it and target the next one, or retry it a week later. Starting at the end of January, Blizzard reduced the damage of all DS monsters in normal and heroic difficulty by 5% per month until its cap at 30% in July 2012. Although players could opt out of this damage reduction, the API did not provide this information.

2.2 Gameplay motivations

There are several player models available in the literature. The founding Bartle model introduces four MMO player types, but was determined qualitatively from Bartle’s experience rather than quantitatively [4]. The four player types in the Demographic Game Design (DGD) model are determined quantitatively, but do not particularly focus on MMO players, so this model may not be the most accurate and valid [5, 6]. An unsupervised method proposed by Anders et al. uses the location and occurrences of an avatar’s death to cluster players into four types, but this model is specific to the Tomb Raider adventure game [12]. Another unsupervised technique introduced in [13] uses correlation networks to predict the achievements a player will complete based on her past data. While the game analyzed is WoW, the technique only takes into account the achievements of a player, not her interest in other game activities such as managing a guild or trading on the auction house.

Finally, using an unsupervised clustering approach, Yee proposed in [21] three MMO player motivation scores: achievement, immersion, and socializing. We pick this model because it takes into account most of the activities expected to take place in an MMO. For each of the three motivation scores, five 5-point Likert scale questions are given to the respondent to answer. These fifteen questions measure the respondent’s interest in chatting, leveling up, or collecting things for example. We then normalize each of the three motivation scores (mean of 0, standard deviation of 1).

2.3 Protocol and data

The last two authors hosted an online questionnaire and posted links to it on popular WoW and online gaming websites from China and the US in October and November 2011, right before patch 4.3 launched. More specifically, we provided a link to take our survey in simplified Chinese on Chinese gaming websites, while US gaming websites were given a link to the survey in English. Participants who took the survey in Chinese may come from Hong-Kong, China, or Macau, but for simplicity, we call them the CN respondents. Similarly, the English-speaking respondents are the US respondents. In the questionnaire, we ask respondents demographic questions such as their age and gender, as well as WoW-specific questions such as the number of years they

have been playing the game for, and whether they ever stopped playing the game for at least a month. Yee’s 15 gameplay motivation questions give us the achievement, immersion, and social scores for each participant. We exclude participants who took the survey after November 30, 2011. In the end, our data consists of 1350 players, among which 29% are women, and 41% from the CN region. The average CN respondent is 23 years old, while the average US respondent is 35 years old. Both the average CN and US respondent have been playing WoW for close to 5 years as of November 2011. We note that the players who frequent the websites on which we advertised may be more expert and dedicated than the average player. However, our sample matches the samples found in previous works in terms of age and gender statistics [20, 21].

In the questionnaire, we also ask respondents the names and servers of their active characters. This allows us to match their survey data with data generated by their game characters. Character data is exposed by the Armory, a Blizzard web service connected to the WoW database¹. Character data was pulled every day between December 1, 2011 and June 30, 2012. Overall, 4389 characters logged in at least once during the 7 months. Distinguishing by character level, 65% of all characters were already level 85 as of December 1st, 2011, 7% reached 85 sometime between December and June, and 28% never reached 85. In other words, two thirds of the active characters are in the endgame phase and able to raid FL and DS.

While the API provides hundreds of measurements at the character level, we focus on three player-level monthly aggregates. First, we measure the number of characters used by each player during the month. In the rest of the paper, we refer to the *active player base* of a particular month as the sample of players who used one or more characters during that month. We refer to the *total player base* as the 1350 players. Second, we define the number of raids that the player entered during the month as the sum of all the non-LFG 5-, 10-, and 25-player raids entered by all her characters that month. The Armory API did not provide the number of “Looking for Group” raids entered. And finally, we define the number of PvP occurrences that the player participated in during the month as the sum of all the duels, arenas, and battlegrounds that the characters of the player participated in. Zone-wide PvP battles such as Wintergrasp are not returned by the Armory API. Table 1 summarizes the 11 features used in this paper.

Domain	Features
Demographic	age, gender, region
WoW-specific	WoW years, stopped before
MMO motivations	achievement, immersion, social
Monthly game data	characters played, raids, PvPs

Table 1: List of all player-level features.

Table 2 provides the average number of characters played over the seven months across gender and region. It also breaks down the total number of raids and the total number

¹The documentation of the Armory API lives at <http://blizzard.github.io/api-wow-docs/>

of PvPs participated in over the seven months. The average CN player uses fewer characters, but raids as much, and PvPs three times more often, than the average US player. CN players seem more competitive and more focused on a single character than US players. Compared to the regional differences, the gender differences are minimal.

	CN		US	
	women	men	women	men
n	84	469	313	484
characters played	1.5	1.3	3.0	2.5
raids (total)	28.6	26.3	23.6	20.2
raids 5	18.3	16.5	16.2	13.1
raids 10	4.1	3.6	3.0	2.4
raids 25	6.1	6.2	4.5	4.7
PvPs (total)	12.9	14.6	4.0	4.9
duels	2.1	4.2	0.4	0.5
arenas	8.9	8.1	2.2	2.6
battlegrounds 10	0.5	0.5	0.2	0.1
battlegrounds 40	1.4	1.9	1.2	1.8

Table 2: Stats of the average player over 7 months, split by region and gender.

3. PRESENCE

For a given month, we define the *ratio of active players* as the size of the active player base divided by the size of the total player base. In this section, we focus on the speed at which of the ratio of active players decreases over the seven months.

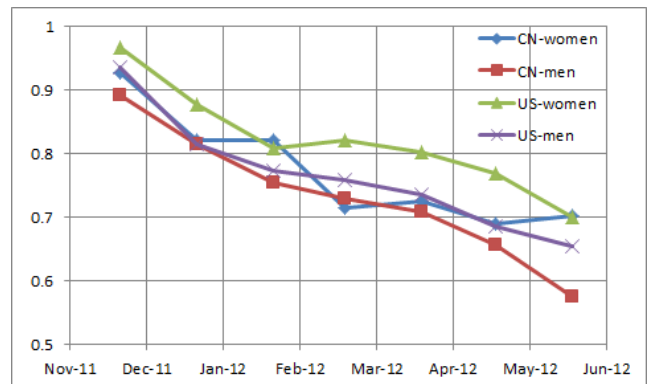


Figure 1: Ratio of active players, split by region and gender.

In December 2011, the active player base represents 93% of the total player base, ie the ratio of active players is 93%. This percentage decreases gradually to reach 64% in June 2012. In other words, a third of the player base stops playing after seven months. Figure 1 plots the decline of the ratio of active players across region and gender. CN men seem to churn twice as fast as US women, with CN women and US men somewhere in-between. We distinguish three overall phases: first a loss of 15% from December to February (5% churn per month), then somewhat flat until April (2% per month), and a final drop of 10% until June (again 5% per month). While the first drop may reflect the speed at which expert players are done with new game content and leave

the game, the final drop may only reflect the seasonal effect of the Summer holidays.

Around 82% of the total player base report having ever taken a break of at least one month before November 2011. These players churn at nearly 5% per month. On the other hand, the players who report having never taken a break before churn at 2% per month on average. Figure 2 illustrates these numbers.

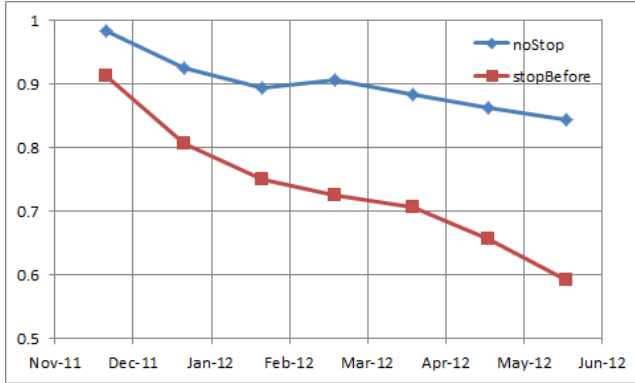


Figure 2: Ratio of active players, split by “having taken a break before November 2011”.

Summary: CN players churn faster than US players. Gender and age make no difference. 15% of the player base churns in the first three months.

4. PROGRESSION AND CONSUMPTION

How does the average player in December differ from her June counterpart? We must look at the progression over the months of the numbers of characters played, PvPs, and raids. However, we saw that there are more and more inactive players over time. These inactive players play zero character, and participate in zero PvPs and raids. For this reason, the metrics of the average player would be much lower in June than in December. That is why in this section, we focus solely on the active player base of each month. In other words, a player only contributes to a month if he played at least one character that month.

The number of characters played by the average active player stays relatively flat: between December and June, it drops from 1.8 to 1.6 for the average active CN player, and from 3.6 to 3.2 for the average active US player. The number of PvPs decreases from 23 to 16 PvPs for the average active CN player, and from 7.6 to 3.5 for the average US player. The number of raids decreases even more steeply. The average active CN player raided 55 times in December, but only 22 times in June. The US counterpart 44 and 17. Between December and June, players use as many characters, yet participate in half as many PvPs and raids. To help us explain this phenomenon, we take a deeper look at the raiding behavior over the seven months.

4.1 Dragon Soul raiding

Dragon Soul contains eight boss monsters to be defeated sequentially. Morchoc is the first and easiest, Ultraxion fifth, and Madness of Deathwing last and most difficult. Figure 3

plots the ratio of active CN and US players who killed these monsters at least once for each month.

In both regions in December, a third of the active player base manages to kill Morchok, 20% Ultraxion, and 7% Madness in normal difficulty. These percentages may seem low, but the Armory API only returns the number of bosses defeated, not the total number of times the player attempted to defeat a boss (and eventually failed). Previous work conducted during a previous expansion of WoW reports that a group of players fails 10 times on average before being able to defeat a boss [3]. Thus the number of players who actually tried defeating normal difficulty bosses may be much higher.

Moreover, the ratio of players defeating a particular boss first increases, then peaks, and finally falls in both regions. In heroic difficulty, these peaks may happen after June. Unsurprisingly, the more difficult the boss, the later it peaks. For example, in normal difficulty, Morchok peaks in January, Ultraxion in March, and Madness in April.

We identify three major differences between US and CN players. First, all normal bosses, as well as heroic Morchok, peak around 40% in CN. For US players, the peaks decrease with the boss difficulty (or with time): 40% for normal Morchok, 35% for normal Ultraxion, 30% for normal Madness, and 23% for heroic Morchok. Second, active US players defeat more and harder bosses in December than CN players. Moreover, the slope of a boss indicates how fast the active player base manages to defeat it. These slopes are steeper for CN than for US players. Thus US players start better, but progress slower. And finally, normal Morchok drops abruptly after March for CN players, but more smoothly for US. In June, nearly 25% of US players defeat it, versus only 12% of CN players. Since Blizzard reduces the damage of DS monsters by 5% every month starting in January, difficulty should not be an issue. It could be that US players find more replay value in re-defeating easy bosses than difficult ones.

4.2 Firelands raiding

FL in patch 4.2 is DS in patch 4.3. Looking at the raiding behavior in FL tells us how much replayability a dungeon has in the eyes of the players. Players may raid FL for gold, equipment (even though DS bosses drop better equipment), to gain an achievement, or just for fun. With the better equipment dropped in DS and sold in the auction house, FL should be easier in December than in November. Therefore, and as shown in Figure 4, it is not surprising that around 30% of CN and US players manage to defeat normal Baleroc in December. In both regions, fewer and fewer players defeat normal FL bosses over time.

The ratio of players defeating heroic Alysrazor and Baleroc rises from 5 to 10% for CN players, but stays flat at 3% for US players. Similarly, the ratio of players defeating heroic Ragnaros rises from 0 to 5% for CN players, but remains at 0% for US players. These findings complement what we previously found for DS: CN players find more replay value in challenging raids than US players.

Summary: Nearly half of the active player base raid every month. The number of raids entered by active players is

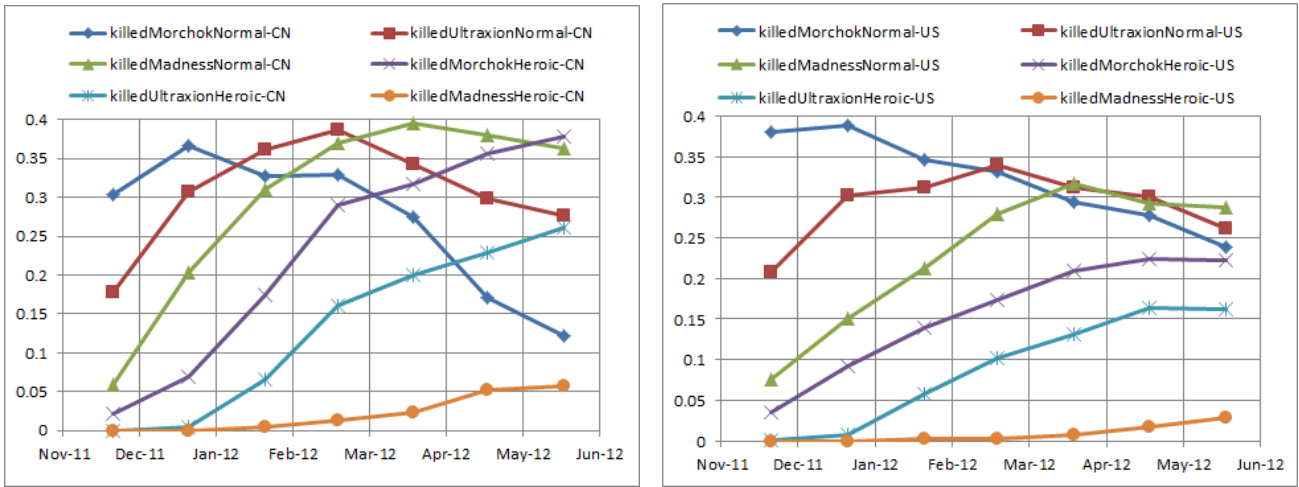


Figure 3: Ratio of active player base who killed a particular boss in the Dragon Soul dungeon, split by region.

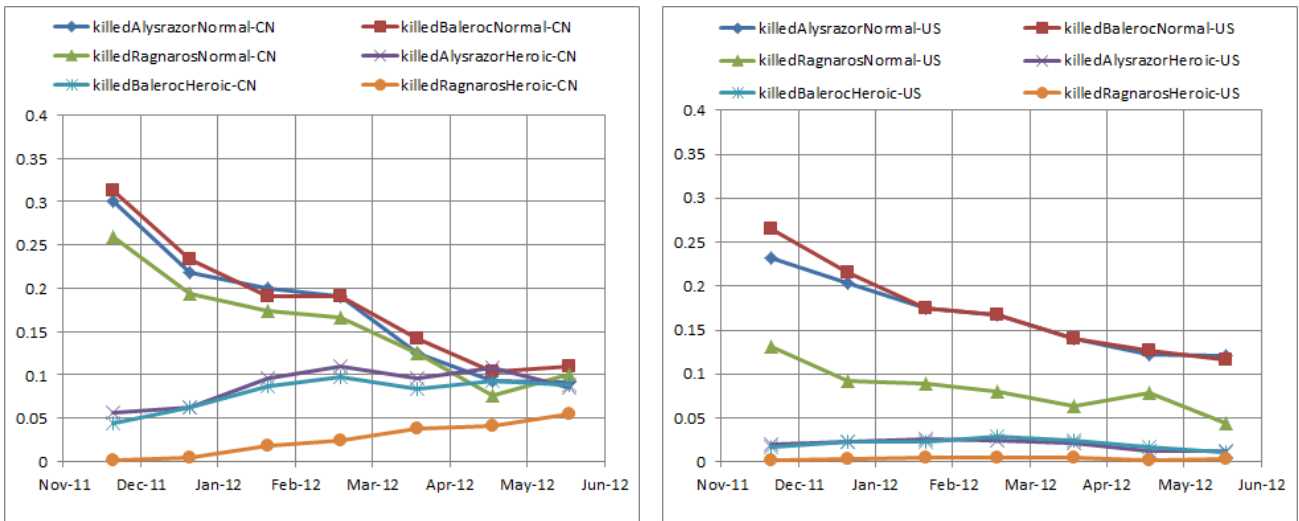


Figure 4: Ratio of active player base who killed a particular boss in the Firelands dungeon, split by region.

feature	January Presence		May Presence		January Churn		May Churn	
	log OR	p	log OR	p	log OR	p	log OR	p
characters played	.44	<.001	.87	<.001	-.25	<.001	-.30	<.001
number of raids	.05	<.001	.10	<.001	-.05	<.001	-.09	<.001
region is US	-.36	.06	-.28	.08	.77	.001	-	>.1
stopped before	-	>.1	-.50	.04	-	>.1	1.10	.004
PvP occurrences	-	>.1	.025	.07	-	>.1	-	>.1
achievement score	-	>.1	-.14	.08	-	>.1	.18	.10
age	-	>.1	-	>.1	-.05	<.001	-	>.1
other 4 features	-	>.1	-	>.1	-	>.1	-	>.1

Table 3: Log odds ratios output from the logistic regression models fitting demographic and game data to presence and churn of January and May 2012. Log odds ratios with p-values above .1 are insignificant, and therefore represented using -.

halved in seven months. CN players take on more challenging bosses than US players.

5. PREDICTING PRESENCE AND CHURN

MMO companies may want to predict whether a player will play next month. The naive approach we first took, detailed below, focuses on predicting presence: will a player login next month or not. Then we realized that predicting presence is relatively trivial. It is more challenging to predict churn: among the active player base, which players will leave the game?

5.1 Predicting presence

We have 8 demographic, WoW-specific, and motivation features about each player, as well as 3 game data aggregates for each month and each player. Can we predict from these features if a player will play at all a couple months after the release of patch 4.3? Does this model still apply six months after the release?

The first model we propose is the simplest: a player will play next month if she played this month. This model is obviously of limited practical use, but provides a baseline for the performance of more complex models. We put ourselves in the place of Blizzard in January 2012. Applying this simple model using the January data of the total player base to predict the February presence gives a precision of 0.89 and a recall of 0.95. These precision and recall set the bar very high for our next model. Using the May data to predict presence in June gives similarly high precision and recall. These precisions and recalls are summarized in table 4.

Prediction	Model	Precision	Recall
Feb. Presence	Same as Jan.	.89	.95
	Regression	.91	.93
June Presence	Same as May	.85	.92
	Regression	.90	.87
Feb. Churn	Regression	.28	.90
June Churn	Regression	.19	.89

Table 4: Precision and recall for the baseline and regression models predicting February and June presence and churn.

The second and more complex model is a step-wise multivariate logistic regression. The regression takes as input a train set containing the eleven features of December (the past, as of January 2012) about each player. January presence is the label to predict (the present). For each feature, the regression outputs a log of the odds ratio (log OR) for two groups differing in one unit of that feature, and a significance level. If the log OR is positive, it means the feature is positive correlated with presence the next month. Logistic regression has been used in previous literature on customer relationship management to predict churn, and showed results comparable to other classifiers such as random forests or neural networks [8].

The resulting model is described in the left-most column of table 3. Only three features are significant and kept in the model. The first feature is the number of characters played in December: the odds of playing in January are

$\exp(.44) - 1 = 55\%$ higher per character played in December. Put more simply, the more characters played in December, the more likely to play in January, independently of the ten other variables. Similarly, more raiding in December means more chance to play in January. And US players are $\exp(-.36) - 1 = 30\%$ less likely to play in January than CN players. None of the remaining eight features are significant, and are therefore excluded from the model.

We then proceed to test our model and predict February presence. The test set consists of the eleven January features (the present), and February presence as the label (the future). When given the data of a particular player as an input, the model will output a number between 0 and 1. The closer to 1, the more confident the model is that the player will play the following month. We kept the cutoff at the default value of 0.5: the model considers that a player with a value output of 0.5 or above will play the following month. This regression model obtains 0.91 precision and 0.93 recall. Even though it is more complex, it is not strikingly better than the simpler baseline.

We reproduce this process placing ourselves in May. We train a model in a similar fashion using April data and May presence. We test the model using May data and June presence. Like in the January model, the number of characters played, the number of raids, and the region are significant and with log odd ratios of the same signs. Three more features become marginally significant: players who have participated in more PvPs are more likely to play in May, while players who have stopped before or are motivated by achievement are less likely to play in May. The precisions and recalls of these June presence models are similar to those of February.

5.2 Transfer behaviors

To understand why the regression model gains nothing compared to the baseline model, we need to break down what precision and recall mean in the baseline model. When predicting February presence in January, each player can fit in only one of four *transfer behaviors*: either 1) they play in January and will play in February (they will “stay in”), or 2) they play in January and will not play in February (they will “churn”), or 3) they did not play in January and will play in February (they will “come back” to play), or 4) they did not play in January and will not play in February (they will “stay out”). These definitions help us explain the great performance of the baseline model. A precision of 0.89 means that 11% of the total player base churned from January to February. A recall of 0.95 means that 5% of the total player base came back.

For all months, the percentage of players who churn or come back is small compared to the proportion who stay in or stay out, as shown in Figure 5. In January for example, the baseline was right for $100 - 11 - 5 = 84\%$ of the total player base. That is how we obtain high precision and recall using a very dumb model. Note, however, that for all months, the percentage of players who churn is roughly twice the percentage of players who come back. Since the average player in our sample has been playing for nearly five years, the ratio of newcomers seems small, and the game is therefore “leaking” players every month. This leak means that more and more

players are staying out.

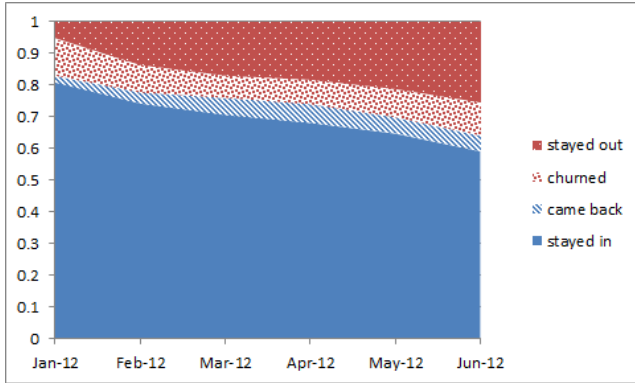


Figure 5: Month-to-month transfer behavior of the whole player base.

5.3 Predicting churn

Predicting presence is not such an interesting problem after all. Instead, we focus on predicting who will churn among the *active* population. Such a model may be more actionable to game companies: they can focus some of their attention on churners to identify potential flaws in their design, or populations that their game is not targeting well.

The baseline model mentioned earlier does not make any sense anymore. It predicts all active players to stay in the game, thereby achieving a recall of 0. We abandon that model. We rebuild the regression models in the same fashion as for presence, except that we only use data from the active player base for a given month, and that the label is whether a player will churn the following month.

As shown in Table 3, the January model for churn is the opposite of the January model for presence: US players are more likely to churn in January, and the more characters played and raids in December, the less likely to churn in January. However, age has become significant: older players are less likely to churn independently of all other variables. Similarly, the May model for churn is the opposite of the May model of presence, except the number of PvP occurrences and region have become insignificant.

After manual tuning using the December-January data, we set the cutoff to 0.1. We obtain very modest precisions of 0.28 and 0.19 to predict February and June churn respectively, but recalls remain around 0.90. This means the models miss out on very few churners, but also flag as churners lots of players who actually stay in. Although their precision is poor, these regression models are simple: they focus on only four features, all obtainable in-game except age. These models provide a good basis for future more complex models.

Summary: Predicting churn is more difficult than predicting presence. Every month, for two players who churn, one is coming back.

6. DISCUSSION

Region has a noticeable impact on retention: CN players churn faster than US players. Looking at progression, and

raiding in particular, active CN players defeat steadily stronger bosses over time. US players, on the other hand, seem content defeating easier bosses, and are less interested in the more difficult ones. CN players also use 1-2 characters on average, compared to 3-4 for US players. CN players seem more focused, and US players more slack. While both populations churn, they do it in their own way. For CN players, presence decreases strongly, not content consumption. For US players, both presence *and* content consumption decrease, but moderately. So when Blizzard’s CEO attributes churn mostly to “the East” [15], it is partially true: activity also decreases for US players.

Other demographic features such as age or gender do not seem relevant to retention. This is somewhat good news for game companies: while a player’s region can be obtained from her IP and the language she plays the game in, gender and age data are not readily available to the game company. Interestingly, gameplay motivations are not significantly correlated with retention.

The number of years playing the game does not improve the prediction of churn. This result confirms a previous study that found little to no correlation between retention and the number of years played [11]. Previous literature on customer relationship management also indicates that the length of customer relationship does not improve models of churn [8]. We expand on previous findings by noting that while players who took a break before November are not more likely to churn by January, they are more likely to churn by May. This tells us two things about MMO players. First, that players who took a break before will very likely do it again. Therefore, MMO companies should focus on anticipating, preventing, and analyzing the first time a player churns, rather than the second or third.

Second, it also tells us that the lifespan of patch 4.3’s content was less than six months for a lot of players. Around the release of patch 4.2, Blizzard’s CEO attributed churn to the lack of new content. Yet Blizzard did not release any new content for nine months after patch 4.3. In expansion 5.0, they released four patches with new content at intervals of roughly three months. Yet the number of subscribers still fell from 10 to 7.6 million between October 2012 and 2013. If lack of content is not to blame, what could be? WoW’s lead game designer believes that players burn out because the game has trained them to complete quests as quickly as possible, rather than taking the time to explore the world [10].

It seems that MMOs need more than just new content to retain players. What could MMO designers do? A naive solution would be to make MMOs more challenging to CN players and easier to US players. Player behavior researchers have developed techniques to assess whether the difficulty of a game situation in a First-Person Shooter game is appropriate [9] to the player’s skill, or if a player’s character loses enough life during a boss encounter [2]. MMO designers could apply such techniques to better tailor raid encounters to their audience. However, tweaking the difficulty on a regional scale is out of question for games with worldwide competition such as guild and PvP rankings. This poses an interesting localization problem: can the same content be

made easy enough to not frustrate US players, yet challenging enough to not bore CN players?

In a given month, the total player base consists of four categories of players: those who stay in, stay out, churn, and come back. Since most people stay in or stay out, presence is not a behavior relevant to predict. Rather, we built a simple regression model predicting the 10% of the total player base who churn every month. This model achieves very poor precision, but it may be improved by adding more in-game data that was not collected, such as the number of words typed by a player during the month, the profits made in the auction house, or real-life information such as a sudden change in lifestyle. Building a model to predict the 5% who come back every month seems even trickier because the features triggering the comeback are likely to be absent from the game. But it is an important question, since for two players who churn, one is coming back. This question may be more easily answered qualitatively.

7. ACKNOWLEDGEMENTS

This research was sponsored by the Air Force Research Laboratory. Thomas Debeauvais performed this work during an internship at PARC.

8. REFERENCES

- [1] M. Ambinder. Valve's approach to playtesting. In *GDC*, 2009.
- [2] M. Ashton and C. Verbrugge. Measuring cooperative gameplay pacing in world of warcraft. In *Proceedings of the 6th International Conference on Foundations of Digital Games*, FDG '11, pages 77–83, New York, NY, USA, 2011. ACM.
- [3] J. Bardzell, J. Nichols, T. Pace, and S. Bardzell. Come meet me at Ulduar: progression raiding in World of Warcraft. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, pages 603–612, 2012.
- [4] R. Bartle. Hearts, clubs, diamonds, spades: players who suit MUDs. *The Journal of Virtual Environments*, 1996.
- [5] C. Bateman and R. Boon. *Twenty-first century game design*. Charles River Media, Inc., August 2005.
- [6] C. Bateman, R. Lowenhaupt, and L. Nacke. Player typology in theory and practice. In *Think Design Play: The fifth international conference of the Digital Research Association (DIGRA)*, DiGRA'11, 2011.
- [7] Z. Borbora, J. Srivastava, K.-W. Hsu, and D. Williams. Churn prediction in mmorpgs using player motivation theories and an ensemble approach. In *2011 IEEE third international conference on social computing (socialcom)*, pages 157–164, October 2011.
- [8] W. Buckinx and D. V. den Poel. Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, 164(1):252–268, 2005.
- [9] A. Canossa, A. Drachen, and J. R. M. Sørensen. Arrrrghh!!!: Blending quantitative and qualitative methods to detect player frustration. In *Proceedings of the 6th International Conference on Foundations of Digital Games*, FDG '11, pages 61–68, New York, NY, USA, 2011. ACM.
- [10] T. Chilton. Blizzard talks the mistakes of training players to hunt for quest markers. <http://www.pcgamesn.com/wow/blizzard-talks-mistakes-training-players-hunt-quest-markers-and-how-pet-battles-won-hearts-community>, August 2013.
- [11] T. Debeauvais, B. Nardi, D. Schiano, N. Yee, and N. Ducheneaut. If you build it they might stay: Retention mechanisms in World of Warcraft. In *Proceedings of the Sixth International Conference on the Foundations of Digital Games*, FDG '11, 2011.
- [12] A. Drachen, A. Canossa, and G. N. Yannakakis. Player modeling using self-organization in tomb raider: underworld. In *Proceedings of the 5th international conference on Computational Intelligence and Games*, CIG'09, 2009.
- [13] B. Harrison and D. L. Roberts. Using sequential observations to model and predict player behavior. In *Proceedings of the 6th International Conference on Foundations of Digital Games*, FDG '11, 2011.
- [14] J. Kawale, A. Pal, and J. Srivastava. Churn prediction in mmorpgs: A social influence based approach. In *Proceedings of the 2009 International Conference on Computational Science and Engineering - Volume 04*, CSE '09, pages 423–428, Washington, DC, USA, 2009. IEEE Computer Society.
- [15] M. Morhaime. Earnings call transcript of q3 2011 results for Activision Blizzard. <http://seekingalpha.com/article/306409-activision-blizzards-ceo-discusses-q3-2011-results-earnings-call-transcript>, November 2011.
- [16] M. Morhaime. World of Warcraft expansions coming on 'a faster cadence,' next already in the works. <http://seekingalpha.com/article/776601-activision-blizzard-management-discusses-q2-2012-results-earnings-call-transcript?page=6>, August 2012.
- [17] E. O'Rourke, E. Butler, Y.-E. Liu, C. Ballweber, and Z. Popovic. The effects of age on player behavior in educational games. FDG '13, 2013.
- [18] R. Pardo. AGC: Blizzard's Pardo on WoW's success. http://www.gamasutra.com/php-bin/news_index.php?story=10773#, 2006.
- [19] B. Phillips. Player experience panel. In *GDC*, 2010.
- [20] D. Williams, M. Consalvo, S. Caplan, and N. Yee. Looking for gender: Gender roles and behaviors among online gamers. *Journal of Communication*, 59(4):700–725, 2009.
- [21] N. Yee. The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. *Presence: Teleoperators and Virtual Environments*, 15, 2006.
- [22] G. Zeschuk. DICE keynote panel. DICE, 2011.