Demographic Profiling from MMOG Gameplay

Peter Likarish[1], Oliver Brdiczka[2], Nicholas Yee[2], Nicholas Ducheneaut[2], and Les Nelson[2]

[1] Dept of Computer Science
The University of Iowa
Iowa City, IA 52245
[2] Palo Alto Research Corporation
3333 Coyote Hill Rd
Palo Alto, CA 94304

Abstract. This paper examines profiling basic demographic information (gender and age) from the gameplay of 1040 World of Warcraft (WoW) players. The authors develop two monitoring systems to track the players, one based on in-game observation and the other on a data source provided by the operators of the game. We describe and extract four feature sets, each from different assumptions regarding the type and amount of data available to an adversary: 1) a one-time snapshot of each character, 2) a series of snapshots from which we extract features for character progression, 3) a mapping of players to characters that allows us to extract higher level features over all the characters belonging to a player and 4) a superset of the previous three sets.
We show that one can predict gender and age (within ±5 years) for 53% of players using machine learning and one can predict gender and age (within ±1 year) for over 11% of participants solely based on the features monitored by our systems.

1 Introduction

Video games continue to increase in popularity, evolving from a niche hobby into a massively popular activity pursued by millions. In 2009, marketing survey group NPD found that 63% of their survey respondents had played a video game in the last six months while only 53% had been to the movies, laying to rest any doubt that video games have achieved widespread appeal[1]. Massively Multiplayer Online Games (MMOG) are one of the fastest growing segments of the video game market. These games allow millions of people to simultaneously play the same game over an internet connection.

This paper examines whether one can profile online gameplayers solely based on how they choose to play a game. Online profiling of this sort has a variety of applications. Knowing a player's demographic characteristics could allow a company to display advertisements that are more likely to be meaningful or interest to an individual. Knowing demographic details about a player may even enable companies to personalize the game world to that player, making the experience more engaging. Profiling is of particular interest to "social gaming" companies whose tend to have a high churn rate and who may not know anything about their players aside from their style of play.

To determine if it is possible to extract demographic characteristics from gameplay, we observe 1040 individuals on the world's most popular MMOG, World of Warcraft [2]. We examine how well one can predict a person's real world (RW) demographic characteristics based on features extracted from their in-game behavior. We show that one can reliably predict a player's gender and age based on the features extracted in this paper, finding that one can predict gender and age (within ± 1 year) for over 11% of our players. With a wider age range (± 5 years), one can predict gender and age for nearly 53% of players.

As a second contribution, we investigate whether knowing the mapping between players and characters (which player plays each character) improves demographic prediction. Many MMOG players choose to play multiple characters, one "main" character and several "alts". Determining the mapping from players to characters is difficult. We use the ground truth mapping from our players and extract additional features that treat all the characters played by a single player as one entity. Statistical testing confirms that the mapping improves predictions of gender but does not improve predictions of age by a statistically significant amount.

We extract hundreds of features from the players' combat, exploration, achievement and social gameplay, divide the features into sets based on the type of observation (one-time character-based, continual progression-based, player-based or a combined superset of the previous three) required to generate each set, and show that one can accurately predict demographic characteristics for the majority of characters using classifiers or regression models from gameplay data. Models trained on our feature sets predict gender with an F-Measure and ROC AUC up to 0.9. SVM-based regression models trained on our feature sets predict age within ± 5.0-5.5 years (Mean Absolute Error from actual age). Before proceeding, we describe Blizzard's World of Warcraft.

1.1 The World of Warcraft

Our participants play what is currently the most popular MMOG in the world, Blizzard's World of Warcraft (WoW).[3] Due to its popularity, we assume most readers are familiar with basic MMOG mechanics and limit the length of our description. WoW has an active subscriber base of at least 11.5 million users (the last time Blizzard acknowledged a subscription figure in 2008[3]. Current subscriber numbers are estimated at 14-15 million users.). WoW is set in the fictional land of Azeroth, where various races battle for survival. Each WoW player creates one or more in-game alter-egos known as a character. The player selects a race aligned with one of two factions, the Horde or the Alliance, each made up of different races (e.g. elves or orcs).

[3] An expansion to World of Warcraft, Cataclysm, will be released after the publication of this paper and make portions of our description outdated.

The character must also select their class: Druid, Hunter, Mage, Paladin, Priest, Rogue, Shaman, Warlock or Warrior. This decision is the first step toward determining if the character is a Tank, Healer, melee DPS, or ranged DPS. Players often create more than one character: one "main" character and several "alts". These "alt" characters allow the player an alternative gaming experience (via race, class or role).

The players earn money and experience by completing quests and killing mobs (non-player controlled characters). When the player gains a set amount of experience, they level up. Player levels are currently capped at 80. Money is used to purchase equipment to improve the character's skills. One can play most of the game's content by oneself but to access the best equipment and most challenging game content, players need to form groups with others. These groups are formalized as guilds. Guilds also provide an in-game social network for players and can range in size from 1 to several hundred members.

In addition to combating mobs, a player may also fight other players (PvP). PvP can happen in a variety of settings, from large-scale fights during raids on the opposing faction or in battlegrounds, to duels and arena combat. Certain servers are designated PvP servers, allowing a player to attack opposing faction members at any time. Other servers are specified as Player vs Environment (PvE), imposing restrictions on PvP.

The rest of the paper is organized as follows. First, we review previous research, followed by a description of our participants and a detailed discussion of the systems that monitor their play. We then discuss extracted features. The features are described in more detail in Section 4. This is followed by an evaluation of how well we predict demographic characteristics from these features. We discuss implications in Section 7 before concluding with future work.

2 Related work

This paper focuses on extracting demographic variables via gameplay profiling and so we highlight related work that involves this sort of prediction. Hu et al attempted to predict demographics such as gender and age by using a Bayesian framework based on webpage click-through data [4]. In a very large study, Singla and Richardson observe that associates in social networks (even friends of friends) tend to have similar interests and personal characteristics and the strength of that relationship is correlated with their level of similarity [5]. ItemSpider, by Tsukamoto et al, is a social network centered around books. The authors found that people with similar characteristics were interested in similar types of books [6].

In 2007, Jones et al studied anonymized query logs and showed that with a series of classifiers one could map queries to gender, age and location of the user. They had a real-world acquaintance of a target user attempt to identify the target in an anonymized data set and found that personal information often enabled identification [7].

Another subset of this work has focused on predicting demographic data based on linguistic features in electronic media, including Herring's study of gender in electronic communication [8] and Koppel's work on determining the gender (and age) of a text's author or of bloggers [9, 10]. Linguistic features were not available to us in this study but could be incorporated to improve performance.

While there has been a large amount of social science and HCI research on online games, there is a dearth of research on demographic profiling from online gameplay. Examples of this type of work include studies by Ducheneaut et al [11], Yee et al[12, 13], Bessiere et al [14], Nardi and Harris [15] and Williams et al [16], among others. In a related vein, Grimes and Bartolacci examined the potential for using Second Life as a platform to teach profiling online behavior [17].

Finally, Nokelainen et al provides an example of how the demographic predictions can be used to personalize their experience by building a Bayesian model of a user based on a questionnaire they fill out [18].

3 Methodology

We established the ground truth for our predictions of gender and age by recruiting a large group of World of Warcraft players and monitoring their gameplay for a period of 6 months from April 5^{th}, 2010 through October 5^{th}, 2010. We now describe our participants and the recruitment process before giving an overview of the two tools developed to monitor their play. Then, Section 4 describes the various features we were able to extract based on our monitoring tools.

3.1 Participant details

We recruited 1,040 WoW participants, 533 participants from the United States, 512 from Hong Kong or Taiwan. The participants were recruited on forums dedicated to WoW, by publishing an article on the proposed study on popular gaming sites (e.g., WoW.com), through word-of-mouth, and via mailing lists collected during previous studies of online gamers. Participants completed a basic demographic survey as well as listing up to 6 WoW characters they were actively playing. Hong Kong and Taiwanese recruitment was done by scholars residing in these countries, with recruitment materials and surveys translated into appropriate dialects of Chinese. 26.25% of the participants were female. Our participants ranged in age from 18-65. The average age of our sample was 27.04 years old with a standard deviation of 8.22 years. Hong Kong/Taiwan participants were more concentrated in their early twenties, with fewer players over thirty. In contrast, over 41.70% of US players were 31 years of age or older.

The 1,040 participants played 3,862 characters during the course of our study, an average of 3.73 characters per participant (stdev=2.15). 2,034 (52.66%) of these characters were aligned with the Alliance. 1,988 (51.48%) of the characters were female. As one would expect from a mature game (WoW is over 6 years old), a large number of the characters in our study, 69%, have reached the highest level possible. Figure 1 plots the age and level distributions.

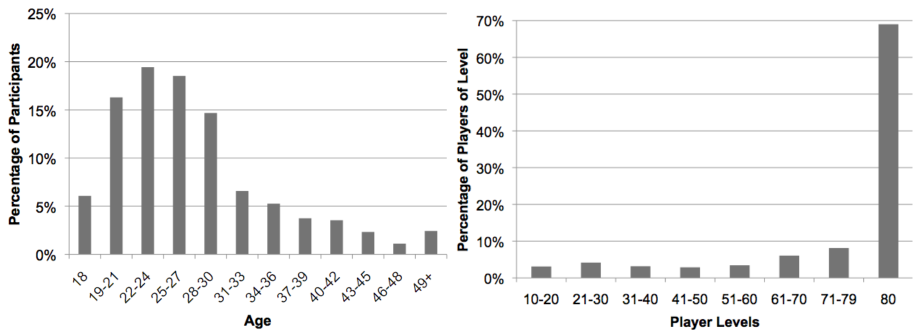

Fig. 1. The age distribution for our participants is plotted on the left and the level distribution for their characters on the right.

3.2 Monitoring participants

The data collection system monitored two sources of data. We conducted in-game monitoring based on a system described by Ducheneaut et al [11]. This system tracked characters on any of the 249 US servers or 31 Hong Kong/Taiwan servers. The software managed 12 WoW robots, each running in a separate virtual machine on one of two Quad Core Mac Pros. The robots log into the game, issued a /who query for the characters they were currently tracking and noted if they were online, collecting in-game location data for each character as well as for their online guildmates. This enables us track with whom our participants are playing, similarly to [12]. The robots cycled through the characters in roughly 45 minute intervals.

A second data collection system consists of a web scraper to gather character information from the WoW Armory [19] in the form of large XML files. The Armory is a Blizzard-provided service that supplies detailed information for all WoW characters over level 10. The Armory includes everything from generic information about a character's race and class to minutiae such as the number of monsters killed, the number and type of deaths and kills, the achievements the character has earned and information about the equipment currently in use. A character's armory entry is updated once per day if that character was active the previous day.

We process the output of both monitoring tools and extract the features described in the next section into a SQL database. A negligible percentage of the data was discarded due to network transfer errors. Heavy load on specific servers had a negligible impact on our in-game collection as well.

4 Feature extraction

Given the data sources detailed in the previous section, we extract a total of 435 features. We divide the features into four levels based on the information required to extract them.

- *Character-based features* are features one could extract given a broad knowledge of a character's gameplay at a single point in time.
- *Progression-based features* are features one could extract given temporal knowledge, that is, a series of character snapshots over a period of time.
- *Player-based features* are features one could extract given character-based features and a mapping between characters and participants/players to enable the extraction of "higher level" features for each participant.
- *Combined features* are a superset of the other three feature sets.

The following sections provide detailed descriptions of the features included in each set. We extract too many features to provide a detailed description of all features in this paper. A list of all our features is available at:
http://www.cs.uiowa.edu/ plikaris/WoWfeats.txt.

4.1 Character-based Features

We extract a set of 246 features for each character from a one-time snapshot of the most recent data for each character taken on October 6^{th}, 2010. We divide the features into categories and summarize the features in each category below.

General features A character's race, class, gender, guild, level, faction, base stats (such as strength or spirit), their professions skill and how they choose to allocate their talent points (after leveling, players can allocate talent points to increase their abilities).

This category also includes miscellaneous information such as how many mounts (rideable NPCs) and pets the character owns, their reputation with various non-player factions and how often they roll greed or need (when a group of characters encounter a valuable object, they "roll" to determine who receives it. The player with the highest roll keeps the object. If a character needs an object, they choose "need", otherwise they select "greed". Need rolls are always higher than "greed" rolls). (total features: 75).

Achievement features WoW grants achievements for completing certain game objectives, such as exploring areas or defeating certain bosses. Achievements categories include: General, Dungeons, Exploring, Feats, Professions, PvP, Quests, Reputation and World Events. We track the number of achievements each character has completed in each of the above categories. We also create a binary complete/incomplete feature for difficult-to-complete specific achievements (total features: 79.

Combat features Combat features include combat-related statistics such as the biggest hit received or dealt by each character. This section also includes the number of deaths the player has experienced (e.g. the number deaths from other players, from NPCs, from falling, from fire, etc) as well as the number of monsters killed. We also track the number of other players killed, and how often the character uses PvP-specific features of the game, such as the arena or battlegrounds. The final piece of information in this category is the value of the character's equipment for each piece of gear they had equipped at the time (total features: 86).

Emotive features We observe the number of times a player hugs, waves, cheers, lols, facepalms, or violins in game. Other emotes are not tracked in the Armory (total features: 6).

4.2 Progression-based features

156 features are extracted from observations of the character's progression over the 6 month period of observation. For each character-based feature that changes over the period of observation, we tracked its rate of change. For instance, the average number of deaths per session played, the increase in equipment value per session played, or the number of hugs per session played. Included in this data set are features such as the percentage of time the character plays on each week day and during which part of the day the character is most active. We omit character-based binary features because completion is a yes/no proposition. An example would be achievements. Progression-based includes the rate of achievement completion in each category but not completion of specific achievements (total features: 114).

The social network data is inherently temporal because it requires multiple observations to establish the network. We calculate a large variety of standard social network analysis metrics including network size, transitivity, centrality, betweenness and clustering metrics for each character. This category also includes information about the variance in racial, class and level balances in the participant's social network (total features: 42).

4.3 Player-based Features

We extracted 33 features at the participant level by analyzing all characters played by a single participant as a group. These are features that are unavailable unless one knows a mapping from players to characters. They include: the percentage of characters of a given gender for each participant and the percentage of characters belonging to each faction for each participant. Other features include the amount of time a participant spends playing each role (melee DPS, ranged DPS, tank and healer).

We also compare the participant's focus in the game relative to other characters. That is, does the player spend more time on PvP or exploring? How much questing do they do relative to other participants? To answer these questions, we divide the participant's into quintiles depending on how many achievements they have completed in comparison to other participants.

4.4 Combined Features

This feature set is a combination of the other three feature sets. We include it to estimate the maximal predictive performance given all information available in this study. To generate Combined, we attached the progression-based features to the character-based features and added the corresponding player-based features for each character.

5 Predicting Real-World Demographics

This section evaluates the accuracy with which one can predict demographic characteristics (gender or age) based on the four feature sets described in the previous section (character-based, progression-based, player-based and Combined).

With the character-based and progression-based feature sets, we make a prediction for each of the 3,826 characters for which we have both character-based and progression-based features. With the player-based feature set, the underlying assumption is that we know the mapping from characters to participants and thus, we make our prediction for each of the 1,040 participants. Combined, a super-set composed of all three feature sets makes a prediction for each of the characters.

We adopt different data mining strategies for predicting gender (a discrete variable) and age (a continuous variable). For gender, we repeatedly trained a C4.5 classifier on each feature set with randomly selected training data. We experimented with other classifiers, including an SVM (Platt's Sequential Minimal Optimization (SMO) [20], an Radial Basis Function (RBF) kernel, varied complexity and gamma), but no classifier substantially outperformed the others. Some previous research binned people into age groups [4]. Instead, we opt to treat age as a continuous variable and to train a regression model to predict age directly. We experimented with several regression models including: linear regression, Partial Least Squares and a Multilayer Perceptron and regression from an SVM model. None outperformed the SVM regression model. We elected to use an SVM regression model [21] to directly estimate each character's (or participant's) age. We used the classifiers as implemented in the Machine Learning Toolkit, Weka [22]

5.1 Description of experiments

We develop two research questions that correspond to the contributions claimed in the introduction:

1. How reliably can we predict gender and age for each character or participant? (RQ1)
2. Which feature set yields the best performance for each demographic characteristic? Does player-based outperform character-based? (RQ2)

To addresses these questions, we carry out two experiments.

Experiment 1: training/test split To address RQ1, we examine the average performance of a model given varying percentages of data reserved for training. We reserve a set amount of data for training a model to predict either gender or age. The amount varies from 10% to 90%. We repeat this evaluation at each percentage split multiple times, randomly selecting the training data to improve robustness. We evaluate gender using F-measure and Receiver Operating Characteristic Area Under the Curve (ROC AUC) and age using Pearson's correlation coefficient, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). This experiment addresses RQ1 by reporting performance metrics of how well we can predict each characteristic given set amounts of training data.

Experiment 2: differences in feature set performance To investigate RQ2, we select a set amount of training data (50% for gender, 80% for age) and increase the number of repetitions with randomly selected training data. We report the differences in performance and perform an ANOVA in order to confirm that there are statistically significant differences in performance. We note that we are able to perform sufficient repetitions that statistically significant differences become the norm rather than the exception.

5.2 Predicting Gender

This section investigates how reliably we can predict gender for each of our feature sets. We address RQ1 (how reliably we can predict gender) and RQ2 (differences in performance between feature sets) using a C4.5 classifier and the two experiment described in the previous section.

Experiment 1: training/testing splits We computed the F-measure and AUC for a C4.5 classifier trained on increasing amounts of data (from 10% of all instances up to 90%). At each percentage split, we randomly selected the training instances each time and using the remaining data to train the classifier, repeating the evaluation twenty times to reduce the effect of random selection on performance. The *a priori* class distribution was preserved in the training and test sets.

Figure 2 plots the results of this experiment. Combined outperformed the individual feature sets when at least 30% of the data was reserved for training. Progression-based underperformed the other feature sets. Character-based and player-based performed roughly similarly according to AUC but player-based dominates character-based when one considers only F-measure.

Classifiers trained on character-based, progression-based and player-based feature sets show limited improvement as the amount of training data increases. This suggests one can predict gender based on the ground truth for a relatively small number of characters. The stable performance with small amounts of training data alleviates concerns that we are overfitting the data.

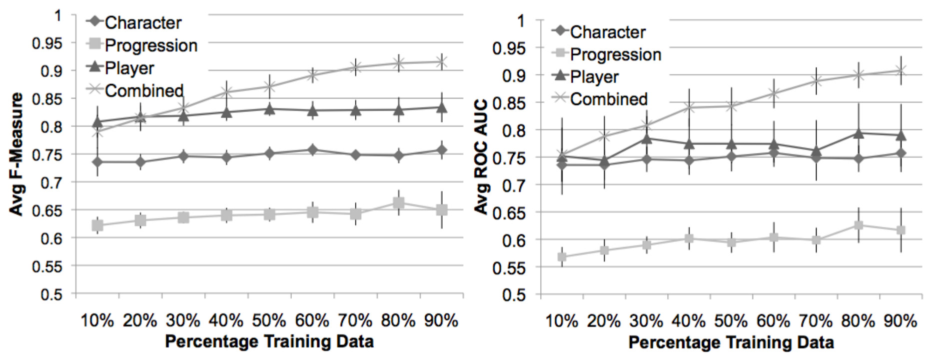

Fig. 2. Average F-measure and AUC of gender for feature sets with varying percentages of data used to train a C4.5 classifier. Training data was randomly selected for each split and the process repeated 20 times.

Experiment 2: differences in feature set performance To ensure we are looking at meaningful differences in performance between feature sets (RQ2), We repeated our random selection of 50% of instances as training data 100 times to be sure of the ordering among feature sets. We preserved the class distribution in both the training and test sets. Table 1 presents the average precision, recall (F-measure is the harmonic mean of precision and recall) and AUC for a C4.5 classifier trained using 50% training data for each of the feature sets and the Combined superset.

The player-based feature set performed better than the character-based and progression-based feature sets. We note that the player-based and character-based AUCs were within approximately standard deviation of one another while the difference in F1 Measure was more pronounced. The progression-based feature set performed poorly, suggesting that in isolation this feature set is the least useful for predicting gender. Of course, we only observed character progression over a six month period and the majority of our characters were mature characters.

Feature set	Precision	Recall	ROC AUC
Combined	0.871 (0.016)	0.871 (0.016)	0.843 (0.028)
Participant	0.826 (0.015)	0.829 (0.014)	0.775 (0.045)
Character	0.775 (0.011)	0.767 (0.011)	0.746 (0.026)
Progression	0.646 (0.015)	0.647 (0.017)	0.602 (0.022)

Table 1. C4.5 results for gender based on 50% training data. Feature sets ordered by decreasing AUC. Standard deviation in parenthesis.

We conducted ANOVAs using either AUC or F-measure as the dependent variable and the feature set type as the factor. The authors emphasize that, due to the large number of observations, statistically significant differences between means of the feature set types are expected rather than remarkable. For AUC, the ANOVA confirmed that the main effect, feature set type, was significant ($F[3, 396] = 1038.074$, $p < 0.001$). Post hoc tests using Tukey HSD showed that the feature sets all differed from one another (means and standard deviations in Table 1, $p < 0.001$). These results suggest that the ordering observed in Table 1 is robust. An ANOVA conducted on F-measure produced very similar results ($F[3, 396] = 4583.297$, $p < 0.001$).

We also examined the most effective features using a traditional feature evaluation metric, Information Gain, to calculate the average rank of features relative to one another across a 10-fold cross validation. Appendix A presents the rank of the top 50 gender features, as well as the set to which each feature belongs. 46% of the top-ranked features belong to the player-based feature set. The difference is even greater for the top 20 features, with 90% belonging to the player-based set. This explains why the player-based set provids better predictions of gender than the character-based set. 42% of the top-ranked features belong to the character-based set but tend to be of lower rank than the player-based features. Only 12% of the features belong to the progression-based set.

5.3 Predicting Age

In this section we use an SVM-based regression model to predict ages. We evaluate RQ1 and RQ2 similarly to gender except that we use a different model (SVM and regression) and metrics used to do so.

We trained an SVM regression model with varying amounts of training data. Figure 3 plots the results of the training/testing split. all four feature sets continued to improve as the amount of training data increased. With regard to RQ1, similarly to our previous results, the progression-based feature set underperformed in comparison to the other sets. The Combined and character-based feature sets both produced correlation coefficients exceeding 0.5, our baseline for a moderately strong correlation.

With 80% of data reserved for training, Combined is able to predict age within ± 5 years in terms of MAE (± 6.5 RMSE), with the character-based feature set able to predict age within ± 5.5 (± 7.5 RMSE) years. Regression models trained with any of the four feature sets perform better than the baseline (the average standard deviation about the mean, depicted as the dotted line in Figure 3) except for progression-based when measured by RMSE. The standard deviation from the average MAE and RMSE varied by over a year. The usefulness of predictions as wide as ± 5 years is discussed in the next section.

Experiment 2: differences in feature set performance Predicting age via regression is difficult (even for with high levels of Combined feature set reserved for training, the MAE of our predictions is ± 5 years from actual ages) and so we reserved 80% of data for training, and used 20% for testing. Table 2 presents the MAE, RMSE or Pearson's correlation coefficient of an SVM regression model

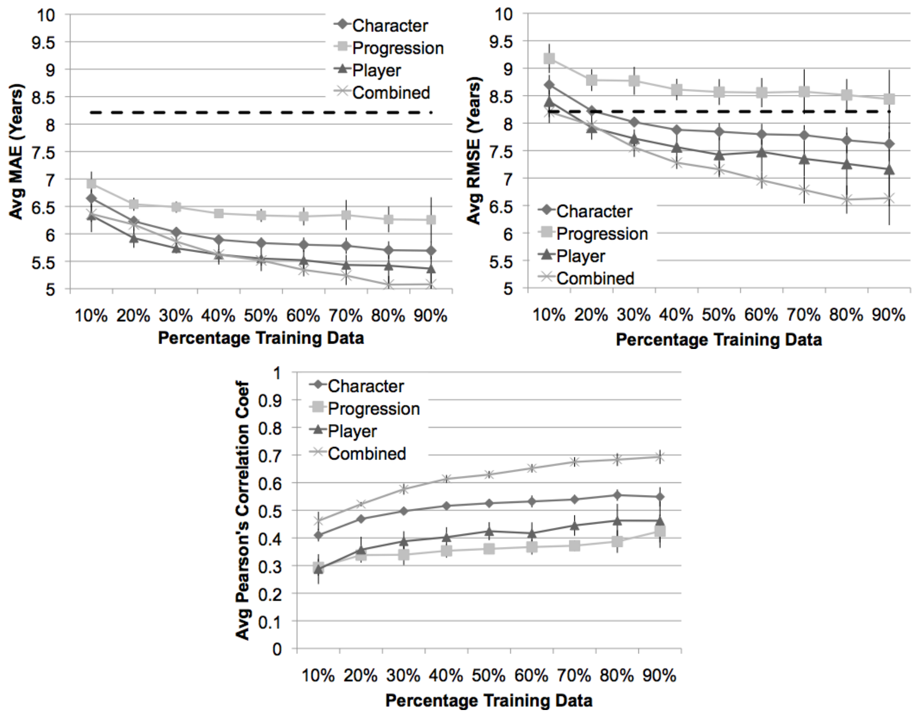

Fig. 3. From top-left, clockwise: MAE, RMSE and Pearson's correlation coefficient of predicted age for each feature set with varying percentages of data used to train an SVM regression model. Lower values represent an improvement in MAE and RMSE. The dotted line is the standard deviation of the distribution about the mean.

trained on each of the feature sets and the Combined superset. We repeated this analysis 25 times.

With regard to RQ2, as was the case with gender, Combined performed substantially better than the other data sets and progression-based substantially worse. The difference in performance between player-based and character-based was well within one standard deviation of both means. While the difference in performance between character-based and player-based was small, the deviation in performance was substantially larger for player-based, with a standard deviation of 2.85 years in terms of RMSE.

We conducted ANOVAs with RMSE and MAE as the dependent variables and feature set type as the factor. The main effect of feature set type was significant for RMSE (F[3, 96] = 105.592, p=$<$ 0.001) and MAE (F[3, 96]= 96.674, p$<$ 0.001). Post-hoc testing using Tukey HSD found that with regard to both RMSE or MAE, there was no significant difference between character-based and player-based (RMSE: $p = 0.110$, MAE: $p = 0.101$). Combined's mean was significantly higher than the other three sets (p$<$ 0.001) and progression-based was significantly lower (p$<$ 0.001). Means and standard deviations are in Table 2.

Feature set	Pearson's	RMSE	MAE
Combined	0.691 (0.009)	6.49 (1.15)	5.01 (0.65)
Participant	0.456 (0.026)	7.35 (2.85)	5.48 (1.64)
Character	0.559 (0.010)	7.60 (1.15)	5.64 (0.751)
Progression	0.388 (0.017)	8.46 (2.02)	6.22 (1.59)

Table 2. Regression model results for age based on 80% training data. Standard deviation in parenthesis. Feature sets ordered by MAE performance.

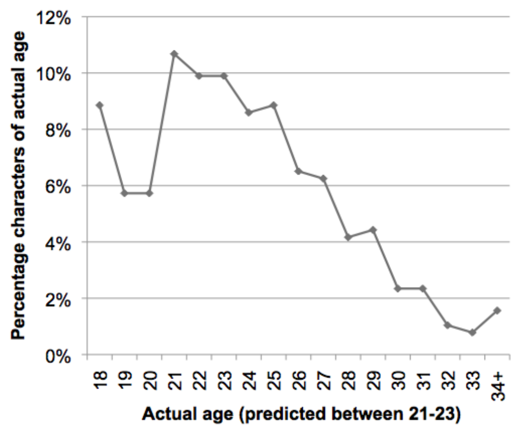

Fig. 4. Probability distribution character is of age given model predicts character is between 21-23 years old. For future predictions within 21-23 age range, we can use this distribution to calculate probability character is of certain age.

Depending on the application, being able to predict a person's age within ±5 years (on average) may or may not be acceptable. We note that in related work, some age bins, particularly for older participants, have been 10 years wide and even larger [4]. It is also possible to treat the prediction generated by a model as a probability distribution rather than an exact value, at least given sufficient amounts of data from previous predictions and their ground truth for a large population of varied ages.

Figure 4 illustrates what such a distribution looks like given that the model has predicted a character is 22 ± 1. The figure was generated from data produced by a model trained on the Combined feature set. We extracted the actual ages of participants for which the model predicted the character was between 21 to 23 years old. The distribution was generated by calculating the fraction of characters that were of each actual age. With this distribution, given that the model predicted a new character is 22, one can determine there is a 48% chance the individual's actual age is 20-24 and only a 7% chance the person is 30+.

As with gender, we ranked the top 50 most predictive features for age using Information Gain (Appendix B). To use the Information Gain algorithm available, we discretized age into three bins. Again, the feature ranking helps to

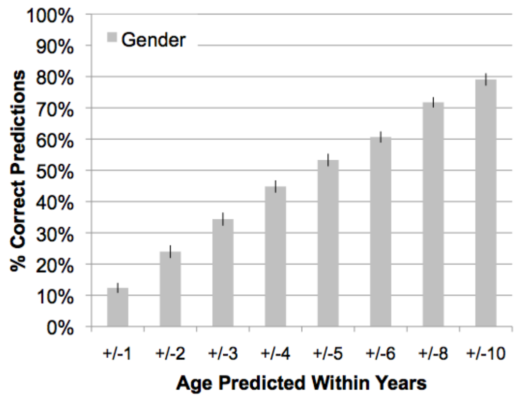

Fig. 5. Percentage of characters with gender correct and age predicted within $\pm x$ years of actual.

explain the relative performance of the feature sets. Unlike with gender (in which player-based outperformed character-based), with age player-based does not outperform character-based. 22% of the top 50 features belonged to the player-based set (25% in top 20). 46% of the top 50 features belonged to character-based and 32% to progression-based.

6 Predicting Multiple Demographic Characteristics

The previous section treated the prediction of gender and age in isolation. However, it could be beneficial to predict demographic characteristics in combination with one another. Making predictions in concert, however, can potentially compound the overall error rates. The purpose of this section is to estimate the percentage of characters for which we can determine both gender and age. Since age is continuous, we calculate the percentage of characters for whom the predicted value falls within a range of $\pm x$ years, where x ranges from one to ten.

We use the Combined feature set and randomly reserve 80% of the instances for training. We then predict a character's age (via SVM regression) and gender (via C4.5 classifier) for the remaining 20% of the data. We repeat this procedure 25 times. The results of this evaluation are presented in Figure 5. For over 11% of the participants, we can predict their gender and age within ± 1 year. For 53% of participants, our models predict gender and age within ± 5 years.

7 Discussion

People typically flock to free services over pay services online. This has limited the business models for companies who operate entirely in a digital space. One popular model is the use of predictive analytics/ad supported services. The techniques in this paper could potentially improve the effectiveness of this business model. Further, if a game developer can determine information about a player based on how they play the game, it may be possible to tailor the game world to better suit that individual, alerting them of gaming events that are likely to be of interest or even individualizing the in-game experience to produce a more engaging experience. Additionally, recommender systems that monitor social gaming and utilize homophily between subjects[4] to improve recommendations could leverage the models generated in this paper to improve recommendations, increasing user satisfaction.

As researchers, we should also explore the implications of rampant monitoring and mining of online activities in order to establish what it is possible to determine from this data. Gaming is no different than any other activity we carry out online: our digital presence constantly leaks information about who we are in the real world. Although not the goal of this paper, profiling can be used to identify who we are even in the absence of personally identifying information. The furor over Blizzard's attempt to tie gaming profiles to their people's "real IDs" reveal that many MMOG players are uncomfortable with efforts to link them to their gaming personas [23].

8 Conclusion

This paper monitors 1,040 online game players and extracts a large number of features based on how the participants play the popular MMOG, World of Warcraft. The high levels of accuracy with which we can predict gender and age from gameplay alone suggest that profiling otherwise anonymous players may allow companies to tailor the gaming experience to individuals. Predictions generated for gender produce an F-measure of 0.75-0.85 with 50% of data reserved for training. A regression model for age with 80% of data reserved for training predicts actual age with MAE of 5.0-5.7 years. We correctly predict gender and age (± 5 years) for 53% of our participants. Features extracted with knowledge of the character to player mapping does improve predictions of gender over character-based features alone. This is not the case for age. One can also restate this finding: a small number of features extracted from a player to character mapping (33 features) produces the same level of predictive performance as a substantially more detailed set of character-based features (289 features).

Future Work We intend to explore the prediction of less obvious demographic variables such as level of education, income or even personality. Finally, one could investigate how difficult it would be to estimate each feature in-game rather than relying on Blizzard's WoW armory. It would also be interesting to develop and study similar feature sets for a different game to explore if these findings are generalizable across games and genres.

---

[4] Roughly defined as the tendency for individuals to associate with individuals who are similar to them.

References

1. NPD: Entertainment trends in america. http://npd.com/lps/Entertainment_Trends2009 (2009)
2. Blizzard Entertainment: World of Warcraft. (http://www.wow.com)
3. Blizzard Entertainment: World of warcraft subscriber base reaches 11.5 million worldwide. (http://us.blizzard.com/en-us/company/press/pressreleases.html?081121)
4. Hu, J., Zeng, H., Li, H., Niu, C., Chen, Z.: Demographic prediction based on user's browsing behavior. Proceedings of the 16th international conference on World Wide Web (2007)
5. Singla, P., Richardson, M.: Yes, there is a correlation:-from social networks to personal behavior on the web. Proceeding of the 17th international conference on World Wide Web (2008)
6. Tsukamoto, T., Nishiyama, H., Ohwada, H.: Itemspider: Social networking service that extracts personal character from individual's book information. LNCS KNOWLEDGE ACQUISITION (2009)
7. Jones, R., Kumar, R., Pang, B., Tomkins, A.: I know what you did last summer: query logs and user privacy. Proceedings of the sixteenth ACM conference on information and knowledge management (2007)
8. Herring, S.: Two variants of an electronic message schema. Pragmatics & beyond. New series (1996)
9. Koppel, M., Argamon, S., Shimoni, A.: Automatically categorizing written texts by author gender. Literary and Linguistic Computing (2002)
10. Schler, J., Koppel, M., Argamon, S.: Effects of age and gender on blogging. 2006 AAAI Spring Symposium (2006)
11. Ducheneaut, N., Yee, N., Nickell, E.: Alone together?: exploring the social dynamics of massively multiplayer online games. In SIGCHI06 (2006)
12. Yee, N., Ducheneaut, N., Nickell, E.: The life and death of online gaming communities: a look at guilds in world of warcraft. In SIGCHI07 (2007)
13. Yee, N., Ducheneaut, N., Nelson, L., Likarish, P.: Introverted elves and conscientious gnomes: The expression of personality in world of warcraft. In SIGCHI11 (2011)
14. Bessiere, K., Seay, A., Kiesler, S.: The ideal elf: Identity exploration in world of warcraft. Cyber Psychology and Behavior (2007) 530–535
15. Nardi, B., Harris, J.: Strangers and friends: Collaborative play in world of warcraft. CSCW06 (2006)
16. Williams, D., Yee, N., Caplan, S.: Who plays, how much, and why? debunking the stereotypical gamer profile. Journal of Computer-Mediated Communication (2008)
17. Grimes, G., Bartolacci, M.: Second life: A virtual world platform for profiling online behavior for network and information security education? an initial investigation. International Journal of Interdisciplinary Telecommunications and Networking (2010) 60–64
18. Nokelainen, P., Tirri, H., Miettinen, M., Silander, T., Kurhila, J.: Optimizing and profiling users online with bayesian probabilistic modeling. Proceedings of the International Networked Learning Conference of Natural and Artifical Intelligence Systems Organization (2002)
19. Blizzard Entertainment: WoW Armory. (http://www.wowarmory.com)
20. Platt, J.C.: Fast training of support vector machines using sequential minimal optimization. Advances in kernel methods: support vector learning (1999) 185–208

21. Smola, A., Schoelkopf, B.: A tutorial on support vector regression. NeuroCOLT2 Technical Report NC2-TR-1998-030 (1998)
22. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.: The weka data mining software: An update. SIGKDD Explorations **11** (2009)
23. M.G.: A blizzard of protest over privacy. (http://www.economist.com/blogs/babbage/2010/07/online_gaming)

A Top 50 features for gender

Rank	Feature	Set	Rank	Feature	Set
1	perc_male_chars	Player	26	defenses_armor	Player
2	fem_male_diff	Player	27	stat_armor	Player
3	num_male	Player	28	isolates	Progression
4	num_female	Player	29	melee_expertise	Player
5	char_gender	Character	30	heal_recd	Character
6	perc_tank	Player	31	arena_played	Player
7	perc_alliance	Player	32	blows_bg	Character
8	fewest_achievs	Player	33	prime_role	Player
9	rel_hugs	Player	34	hon_kills_world	Character
10	melee_main_dps	Character	35	hon_kills_total	Character
11	perc_range_dps	Player	36	total_dmg_dealt	Character
12	respecs	Character	37	duels_won	Character
13	perc_melee_dps	Player	38	equip_epic_items	Character
14	rel_pvp	Player	39	hon_kills_pvp	Character
15	hord_alli_diff	Player	40	central_close	Progression
16	stat_str	Character	41	char_class	Character
17	hit_recd	Player	42	num_hugs	Character
18	achiev_sum_pvp	Player	43	duels_lost	Character
19	total_dmg_recd	Player	44	stat_spi	Character
20	total_heal_recd	Player	45	melee_off_dps	Character
21	most_achievs	Player	46	deaths_raiddung	Character
22	melee_power_base	Character	47	transitivity	Progression
23	t_total_heal_recd	Progression	48	defenses_dodge	Character
24	t_total_dmg_recd	Progression	49	achiev_tab_pvp_total	Character
25	stat_stamina	Player	50	t_duels_won	Progression

Table 3. Most useful features for predictions of gender, as ranked by Information Gain, average rank across 10-fold cross validation.

B Top 50 features for age

Rank	Feature	Set	Rank	Feature	Set
1	perc_melee_dps	Player	26	achiev_tab_pvp_total	Character
2	duels_lost	Character	27	arenas_played	Character
3	duels_won	Character	28	t_dung_lk_25_bosses	Progression
4	perc_bg_wins	Character	29	dung_5play_entered	Character
5	t_duels_won	Progression	30	rel_pvp	Player
6	t_duels_lost	Progression	31	arenas_won	Character
7	need_rolls	Character	32	t_hon_kills_pvp	Progression
8	perc_need	Character	33	arenas_played	Character
9	respecs	Character	34	achiev_sum_pvp	Character
10	t_summons	Progression	35	summons	Character
11	most_achievs	Player	36	avg_lvl	Player
12	num_daysPlayed_month	Player	37	deaths_other_players	Character
13	t_need_rolls	Progression	38	t_blows_bg	Progression
14	fewest_achievs	Player	39	emblems_valor	Character
15	t_hon_kills_total	Progression	40	t_deaths_total	Progression
16	CallCrusade25_PlayerRaid	Character	41	LKDungeon	Character
17	CallCrusade10_PlayerRaid	Character	42	t_dung_lich_25done	Progression
18	prop_80	Player	43	t_deaths_falling	Progression
19	t_total_kills	Progression	44	dung_lich_10play_done	Character
20	greed_need_ratio	Character	45	blows_arena	Character
21	fem_male_diff	Player	46	achiev_tab_pvp_Arena	Character
22	t_total_heal_recd	Progression	47	t_total_dmg_recd	Progression
23	num_female	Player	48	deaths_warsong	Character
24	lvl_var	Player	49	t_lich_25_bosses_killed	Progression
25	t_deaths_other_players	Progression	50	rel_dungeons	Player

Table 4. Most useful features for predictions of age, as ranked by Information Gain, average rank across 10-fold cross validation.