Data-Driven Games User Research

Thomas Zimmermann
Microsoft Corporation
tzimmer@microsoft.com

Bruce Phillips
Microsoft Corporation
bphil@microsoft.com

Nachiappan Nagappan
Microsoft Corporation
tzimmer@microsoft.com

Chuck Harrison
Microsoft Corporation
charlesh@microsoft.com

CHI 2012, May 5–10, 2012, Austin, TX, USA.
ACM 978-1-4503-1016-1/12/05.

Abstract
In this paper we provide a brief introduction to some of the games user research activities at Microsoft. We focus on the analysis of automatically collected game data. We will show how this data can lead to insights about game usage and player progression.

Keywords
Games, Analytics, Game Usage, Player Progression

ACM Classification Keywords
K.8.0 [Personal Computing]: Games.

General Terms
Human Factors, Measurement.

Introduction
Gaming companies now routinely apply data mining to their user data in order to provide a better user experience and to plan the next release of their software. In this position paper we will present some of the games user research activities at Microsoft that leverage such automatically collected data. We will show how this data can lead to insights about game usage and player progression and discuss our current plans for future work.

Data Analysis

In addition to studies with actual gamers, we use a wide variety of data for our research:

In-game data: For some games, we have detailed usage data which is collected through instrumenting the game. Often such data is collected as part of beta tests, which are releases of a nearly-complete version of a game to a limited set of players. Examples of such data are the games modes that are played, but also more fine-grained information such as the coordinates where players died, or how they alternated between weapons.

Xbox Live data: For the Player Experience Panel [2] — a group of approximately 18,000 international players who have volunteered to participate in research with the Games User Research group at Microsoft — we collect game history information: the games played, achievements earned, as well as presence information. We get the data through an XML feed from the Xbox Community Developer Program [3]. In our research, we often use this data to better target surveys.

Online communities: For some projects, we monitor internet forums to better understand and support the quantitative findings in our research. We are currently experimenting with natural language processing (NLP) techniques to automatically extract topics relevant to game experience from online discussions.

Example: Game Usage Data

Project Gotham Racing 4 (PGR4) is an auto racing game for Xbox 360 published in 2007. The game features 100+ vehicles in 7 classes, 100+ routes, and a large number of event types and game modes. In previous work [1], we mined 3.1 million entries of game log data, from several thousands of users who played the game. Each entry contains the type of event, the route, the vehicle used and the number of vehicles in the race as well as the career rating and the number of events they've completed for the players.

As shown in Table 1, the most commonly used *game mode* is a single player mode (OFFLINE_CAREER). The multiplayer modes PGR_ARCADE and NETWORK_PLAYTIME are also used in a significant percentage of races. Other modes are played less frequently with mode NETWORK_TOURNAMENT_ELIM (elimination rounds in a network tournament when racing against a large pool of players) being used in less than 0.1% of races.

Table 1. Usage of game modes in PGR4.

Game Mode	Races	(% of Total)
OFFLINE_CAREER	1479586	(47.63%)
PGR_ARCADE	566705	(18.24%)
NETWORK_PLAYTIME	584201	(18.81%)
SINGLE_PLAYER_PLAYTIME	185415	(5.97%)
…		
NETWORK_TOURNAMENT_ELIM	2713	(0.09%)

These number show that some game modes were not user very often. The underutilization of content is even more pronounced when looking at *event types* (e.g., street race, elimination race, etc.) where 12 of the 29 event types were used in less than 1% of races. In a similar way, out of 134 unique *vehicles*, 50 were used in less than 0.25% of races, and 16 in less than 0.1%.

Each game mode, event type, and vehicle represents significant investment during game development.

Knowing which ones are played the most can help to prioritize resources and ultimately provide a better gaming experience.

Example: Player Progression

By analyzing achievements we can obtain insights into how games are played. Figure 1 shows a visualization called the *player progression chart* (taken from previous work [2]) for a sample of roughly 18,000 players.

Each dot represents a distinct achievement available in Halo 3, across the entire game. It shows achievements for completing single player missions, completing challenges such as finding skulls, for accomplishments in multiplayer, and everything else. The *X axis shows the average number of days it took the players to get the achievement*. Achievements on the left were earned faster than achievements to the right. The *Y axis is the percentage of players getting the achievement.* Achievements located near the top are earned more frequently than achievements near the bottom.

The graph shows four distinct clusters, which mostly correspond to the various types of achievements available in the game. The blue highlight indicates those achievements attained by playing through the campaign. The red highlight shows those achievements for collecting skulls in the campaign. The purple area shows achievements for playing through the campaign meta-game, where players can turn on campaign scoring and skulls to try to get high scores when completing missions. The yellow area shows achievements earned playing downloadable content.

In Figure 1, we have also marked the game completion achievement on the graph. We can see that about 73% of the players got this achievement (game completion rate), and it took, on average, 26 days to get, that is the number of days from when they first started playing to the day they got this achievement.

We can compute similar data and visualizations for most Xbox games. This allows us to compare the experience across games and correlate this information with other data sources, for example Metacritic ratings. In previous work we found that games with higher ratings are more likely to be completed [2].

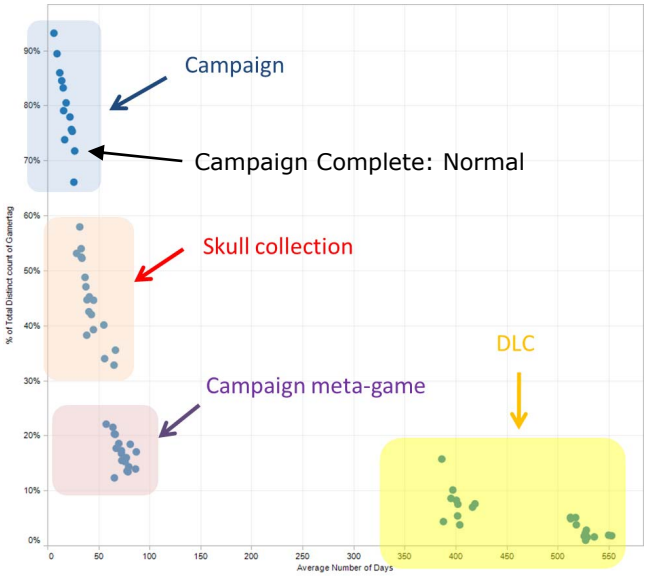

Figure 1. Player progression chart of Halo 3.

Future Work

In the previous sections, we gave examples of mining data to obtain insights into what parts of a game are

used and how players progress within a game. Other aspects for which we mine data are activity, attrition, and tenure within in games. Our current and future work focuses on the following topics:

Social Play: What is the impact of social behavior on player experience in games?

Avatars: How do players personalize their avatar and how do the avatars affect their game experience?

Free-to-play: What is the user experience for games that are based on the free-to-play principle?

For more information about our work, please visit http://research.microsoft.com/en-us/groups/ese/ and http://mgsuserresearch.com/ Thanks!

Biographies

Thomas Zimmermann is a researcher in the Research in Software Engineering Group at Microsoft Research. His research interests include empirical software engineering, computer games, recommender systems, and social networking. He is best known for his work on systematic mining of version archives and bug databases to conduct empirical studies and to build tools to support developers and managers.

Bruce Philips joined the Microsoft Game Studios User Research group in 2001 and has worked on a wide variety of games, including Mech Assault, Voodoo Vince, and Halo: Reach. Bruce recently moved into a full-time analyst role helping to define the next generation of analytic-based entertainment research.

Nachiappan Nagappan is a senior researcher at Microsoft Research at Redmond, WA. He works in empirical software engineering and is very interested in applying data analytics to game data from a software engineering perspective.

Chuck Harrison joined the Microsoft Game Studios User Research group in 2005. During his MGS tenure, Chuck has helped ship some of Xbox Live Arcade's most successful titles, some web-based games, and retail hits such as Halo Wars, Alan Wake and Joy Ride for Kinect. Chuck's team is currently working on a wide variety of games from upcoming Windows-based releases, the next round of XBLA hits, major franchises like Gears of War, and some exciting new Kinect titles. Chuck has over 15 years of user research experience ranging from consumer products to back office and business applications. Before joining Microsoft, Chuck worked at several other high tech companies such as Siebel, BMC, Intel, & Netscape (in the early 90's). Over the years, Chuck has presented research at several HCI and game related conferences and has published researched focused articles and book chapters.

References

[1] K. Hullett, N. Nagappan, E. Schuh, and J. Hopson. Data analytics for game development (NIER track). In Proceedings of the 33rd International Conference on Software Engineering (ICSE), 2011, pages 940-943.

[2] B. Phillips. Peering into the Black Box of Player Behavior: The Player Experience Panel at Microsoft Game Studios. Games Developers Conference (GDC), 2010.

[3] Xbox LIVE Community Developer Program (XCDP) http://www.xbox.com/en-US/Community/Developer/home