Press Start Now! How the Video Game Industry is using Big Data and Analytics in Today's Games.

Ambrose M Malagon Cordero
Bellevue University - DSC500-T304 Introduction to Data Science
amalagoncordero@my365.bellevue.edu

Abstract

As software, video games follow a specific development process where the end result is a product that serves as a creative outlet and generates profit. The development process tends to follow these ubiquitous steps, however developments in technology over the years have continuously driven enhancements; chief amongst these technology breakthroughs is Video Game developers incorporating online connectivity into their products. Internet connectivity serves a twofold purpose: to provide enhancements (such as bug fixes post-launch), services (synchronous multiplayer experiences) and gather data from gamers in regards to the games they are playing. For the purposes of this study, I want to look at the how Big Data and Analytics are impacting the traditional development process: looking at the past, the current status quo and where we are headed. I want to see how data gathering and deep-dive analytics allows Developers to not only improve upon the existing product, but also generate new ideas for new experiences for the gaming community.

Keywords

Console; Telemetry; Gameplay; Devices; Game Metrics; Machine Learning; Game Artificial Intelligence.

ACM Classification Keywords

I.2.1 Applications and Expert Systems (H.4, J)

Introduction

Looking at over 60 years of history, the origin of video games can be traced back to the endeavors of scientists in research labs and Universities [1]. The passage of time demonstrates an unprecedented evolution in terms of presentation, gameplay mechanics and distribution that continue to change even as we speak. Distribution being the key aspect as to why games have become so commonplace household. Once relegated to the physical constraints of media and device, advances in technology and the Internet now allow games to transcend these boundaries and reach audiences that were previously unavailable.

Fast forward to today: Video Games have become one of the most popular and profitable forms of entertainment in the 21st Century; one that rivals the likes of Hollywood's awe-inspiring, revenue-generating mainstream movies [2]. The Video Game Industry's success, as the alluring flame, has drawn out the likes of many who want to express their creative endeavors and take their share of the earnings. Too many developers, producing too many games that are all vying for our attention. The question asked today is how does one stay successful in an industry that's defined by franchise names and that is tagged itself with key success factors such as quality and innovation?

Incorporating the concept Internet connectivity as whole, towards games and the devices we play in, has introduced a new key success factor: the players and how they play games. It is one that Video Game Companies are now exploring through the concepts of Big Data and Data Analytics.

The first part of this paper will explore some key statistics regarding Big Data generated by games (and gamers), why it's considered Big Data and how companies are using Big Data. Next, we take a look at the use cases for Data Science in Gaming, looking at the potential, while at the same time, hypothetical efficiencies. We follow up by taking a look at Machine Learning and its applications in gaming. Finally, we take a look at the Data Science process as a whole and compare it against the Video Game development.

How Gameplay translates to Data, Big Data

According to a study conducted by Nielsen (a global data and measurement gathering company) in 2013, gamers ages 13 and up spend an average of 6 hours a week playing games on any given device [3]. In 2013, the total number of payers worldwide totaled about 1.2 billion people [4]. While the stats can be broken down even further by segment to further exemplify the complex variety of the gaming ecosystem, one can infer that the total number of aggregated gamers multiplied by the total number of hours played translates into a staggering amount of data.

Year	Average Hours per week	% Increase over previous year
2011	5.1	-
2012	5.6	9.8%
2013	6.3	12.5%

Figure 1. Total Hours averaged by player by week by year. Numbers provided by Nielsen.

Video game companies use *telemetry* [5], the act of obtaining any source of data obtained over a distance. Game telemetry data can be thought as of as the raw units of data that are derived remotely from somewhere; an example: an installed client submitting data about how a surer interacts with a game, transaction data from an online payments system or bug/error reports. Gamers create massive data streams about everything they do within a game. Game telemetry data that has been stored and ordered; it is possible to transform the data into interpretable measures known as *game metrics*.

During the Strata 2013 presentation "Video Games: The Biggest Big Data Challenge", At the time Electronic Arts CTO Rajat Taneja stated that the first-person shooter Battlefield generated an over 1 Terabyte of data per day by more than 50 different telemetry events [6]. This data was used to optimize and improve on the existing game to continue ensure player engagement; if players abandon the game if the first levels are too difficult or too easy, data was used to find the bottlenecks within the game.

Game	Data Produced	# of Telemetry Events
Battlefield	over 1 TB	over 50
The Simpsons Tapped Out	150 GB	over 120

Figure 2. Data produced by EA games on a daily basis [6].

Monthly game sessions	Active registered users	Gross daily data captured
2.5 Billion	275 million	50 Terabytes

Figure 3. Total Data produced by EA games on a monthly basis [6].

Data in Use – Data Science Case Uses for Video Games

Electronic Arts is just one example of how the gaming industry realized the benefits of using Big Data as a means to drive customer engagement, optimize the gaming experience and find other leads to make more money [7]. This view can be construed as justification for Big Data as a business intelligence. However, the amount of data the gets generated can be used of data science related tasks and machine learning [8]. Understanding the value of data can promote the search for ways to apply data science and it's benefits in the industry [9]in areas such as:

- Game development:
  - Data Science can be used to build models, analyze and identify optimization points to improve gaming models.
- Game monetization:
  - Big Data is used for predictions on player behavior; optimizing games in such ways that players will return and continue to invest in the game.
- Game design:

    - Data can assist with predicting bottlenecks, reasoning and timing. It can also assist with developing new game concepts, storylines and mechanics.
- Object identification:
    - Experienced with Augmented Reality Games such as Pokémon Go. Distinction between forms, objects, obstacles and figures become easier and much faster for the player.
- Visual Effects and graphics:
    - The use of advanced algorithms and techniques such as Photogrammetry (used in the recent releases such as Resident Evil 2 and Devil May Cry 5) allow developers to create engaging visual effects.
- Social and customer analysis:
    - Customer insights regarding their attitude towards brand and feedback to help segment the audience and adjust the product accordingly.

Artificial Intelligence - Machine Learning in Video Games

Artificial Intelligence in Video games can refer to a broad set of algorithms that incorporate techniques from other disciplines such as control theory, robotics, computer graphics and computer science [10]. As such, Gaming AI may not fit the mold of "true AI" as they may not incorporate some of the latter's standard criteria such as computer learning. Gaming AI is not designed to learn how to beat human players, rather its purpose is to enhance the human player's experience [11]. Some of the more common gaming Artificial Intelligence techniques include:

- Finite State Machine (FSM) algorithm, in which a designer generalizes all possible situations that an AI could encounter, and the programs a specific reaction for each situation.
- Monte Carlo Search Tree (MCST) algorithm, which embodies the strategy of using random trials to solve a problem. The algorithm works by considering all possible options against all player actions, then considers all of the player reactions and so on.

While AI designers work very hard to make Non-Player Characters seem intelligent, the later lack on the ability to actually learn. Since most behavior is programed and unable to learn anything from player's actions or inputs. However, there are some notable examples that exemplify the use of Machine Learning. In this paper we take a look at two games in particular: Dragon's Dogma and Black White.

Dragon's Dogma [12], a single player Action-Roleplaying game in which the player has control over one character (the player character – PC) and is allowed a party of up to three additional NPC characters. The player is allowed to create a second companion character (NPC) labelled as a Pawn by the game. During the Pawn-creation process, the player is provided with a set of multiple-choice questions that will affect the Pawn's behavior in and out of combat. What is unique about this system is that players have an effect on their Pawn's behavior even after the creation process. Should a player fight aggressively and recklessly, then eventually your Pawn goes on to do the same. Start using items to heal the Player Character and the Pawn will start to prioritize healing the Player Character. The Pawn also gains experience fighting against enemies – the more it fights a particular type of enemy, chances are it will learn about their weak spots and try to exploit them. The game had an online component, wherein you can hire up two additional pawns created by other players. While you cannot affect another player's Pawn's inclinations and/or change their equipment, they nonetheless can prove to be useful assets if they are trained correctly and know how to fight the enemies you are up against. A Warrior pawn that knows to flank enemies and wait for the right time to strike is certainly going to be better than a Mage who goes straight at an enemy to strike them with their staff.

Figure 4. Dragon's Dogma [12]. Player using the Knowledge chair to reinforce or deter a Pawn's behavior.

Black and White [13], a simulation video game that puts the players in the role of a god, was praised for its use of Artificial Intelligence. Of particular note is the Creature system, an avatar like creature that has strengths and weaknesses that the player "raises". The player teaches the creature to perform tasks, what and when to eat and how to attack or impress enemy villages. Teaching is performed using a reinforcement learning system: if the creature does something the player does not want it to do, it can be discouraged. Lionhead Studios, the game's developer, used Michael Bratman's belief-desire-intention model [14] to simulate the creatures' learning and decision-making processes:

- A creature forms an intention by combining desires, opinions and beliefs.
- Beliefs are attributed to lists that store data about various world objects.
- Desires are the goals the creature wants to fulfill, expressed as simplified perceptron.
- Opinions describe the ways of satisfying a desire using decision trees.
- For each desire, the creature selects the belief with the best opinion, thus forming an intention or goal.

Both games provide a clear example of how Machine Learning can be used to break the mold of existing gaming AI; by allowing the player to ability to teach an NPC to do things that go past the limitations imposed by a script.

Conclusions and Considerations

The Game Development Life Cycle (GLDC) is described by developers as a life cycle of software development with the objective to entertain the end-user [15]. While the model varies by preference and project scope, the GLDC follows a set of core set of universal steps. The use cases for incorporating Data Science and Machine Learning show enticing promise towards finding efficiencies and innovations. While there are some clear examples in areas such as game monetization and visual graphics [16-17], other areas such as game design and development where not as clear. Another point of contention is that while it appears that companies understand the benefits of utilizing the insights gained from Big Data, it is not so clear that they understand the benefits of incorporating Data Science.

On the other side of the spectrum, we have scholars and researchers using games as a tool to research Artificial Intelligence and Machine Learning [18-29]. While the intention of this research is of academic nature, it nevertheless shows the potential implications of using these technologies as part of the gameplay experience. It may take time before some of the breakthroughs achieved in academic pursuit become a part of the console community. However, the endeavors continue on all fronts and one can only imagine the possibilities that lie ahead.

Acknowledgements

My teachers, peers, coworkers and anyone brave enough through read this: Thank you.

References

[1] History.com. Video Game History. August 21, 2018. Retrieved April 18, 2019. https://www.history.com/topics/inventions/history-of-video-games

[2] Angelo M. D'Argenio. Statistically, Video Games are now the most popular and profitable form of entertainment. July 10, 2018. Retrieved April 18, 2019. https://www.gamecrate.com/statistically-video-games-are-now-most-popular-and-profitable-form-entertainment/20087

[3] Nielsen. Multi-Platform Gaming: For the Win. May 27, 2014. Retrieved April 18, 2019. https://www.nielsen.com/us/en/insights/news/2014/multi-platform-gaming-for-the-win.html

[4] Anders Drachen. What is Game Telemetry? August 12, 2012. Retrieved April 18, 2019. https://gameanalytics.com/blog/what-is-game-telemetry.html

[5] Taylor Soper. Study: 1.2 billion people are playing games worldwide; 700M of them are online. GeekWire. November 25, 2013. Retrieved May 13, 2019. https://www.geekwire.com/2013/gaming-report-12-billion-people-playing-games-worldwide/

[6] O'Reilly. Strata 2013: Rajat Taneja, "Video Games: The Biggest Big Data Challenge". Video. February 27, 2013. Retrieved April 20, 2019 from https://www.youtube.com/watch?v=ZK_PXlbvOfM

[7] DataFloq. The Gaming Industry Turns To Big Data To Improve The Gaming Experience (posted October 28, 2018). Retrieved April 4, 2019 .https://datafloq.com/read/gaming-industry-turns-big-data-improve-gaming-expe/137
[8] Graham Annett. How can a data scientist get into game development? Packt.com. August 7, 2017. Retrieved April 21, 2019. https://hub.packtpub.com/how-can-data-scientist-get-game-development/

[9] ActiveWizards. Top 8 Data Science Use Cases in Gaming. Retrieved April 20, 2019. https://activewizards.com/blog/top-8-data-science-use-cases-in-gaming/

[10] Wikipedia. Artificial Intelligence in Video Games. Article. Retrieved May 14, 2019. https://en.wikipedia.org/wiki/Artificial_intelligence_in_video_games.

[11] Harbing Lou. AI in Video Games: Toward a More Intelligent Game. Blog. August 28, 2018. Retrieved May 14, 2019. http://sitn.hms.harvard.edu/flash/2017/ai-video-games-toward-intelligent-game/

[12] Paprika. Pawn Guide (Mechanics/Basic Knowledge). Steam Community discussions. January 9, 2016. Retrieved May 14, 2019. https://steamcommunity.com/app/367500/discussions/0/451850468363360936/

[13] Wikipedia. Black and White (video game). Article. Retrieved May 14, 2019. https://en.wikipedia.org/wiki/Black_%26_White_(video_game)

[14] Academic Dictionaries and Encyclopedias. Belief-Desire-Intention software model. Article. Retrieved May 15, 2019. https://enacademic.com/dic.nsf/enwiki/951131

[15] Juego Studios. What is the Game Development Life Cycle? April 13, 2017. Retrieved April 18, 2019. https://www.gamasutra.com/blogs/JuegoStudios/20170413/296019/What_Is_the_Game_Development_Life_Cycle.php

[16] NvidiaGameWork. Deep Learning for Game Developers. Video. In GDC 2018 presentation on Deep Learning for Game Developers. June 7, 2018. Retrieved April 21, 2019. https://www.youtube.com/watch?v=rF6Usm0tdhk

[17] Meetupvideo United Kingdom. Data Science for Game Analytics at King. In Data Science Festival. July 26, 2016. London, UK. July 28, 2016. Retrieved April 21, 2019. https://www.youtube.com/watch?v=ez-4m2_jRqQ

[18] Paul Bertens, Anna Guitart and África Periáñez. 22-25 August 2017. Games and big data: A scalable multi-dimensional churn prediction model. In 2017 IEEE Conference on Computational Intelligence and Games (CIG). IEEE Explore. New York, NY, USA. 2325-4289. DOI: https://doi.org/10.1109/CIG.2017.8080412

[19] Niels Justesen and Sebastian Risi.. 22-25 August 2017. Learning macromanagement in starcraft from replays using deep learning. In 2017 IEEE Conference on Computational Intelligence and Games (CIG). IEEE Explore. New York, NY, USA. 2325-4289. DOI: https://doi.org/10.1109/CIG.2017.8080430

[20] G.N. Yannakakis, J. Levine and J. Hallam. 19-23 June 2004. An evolutionary approach for interactive computer games. In Proceedins of the 2004 Congress on Evolutionary Computation (IEEE Cat. No.04TH8753). IEEE Explore. Portland, OR, USA. 0-7803-8515-2. DOI: https://doi.org/10.1109/CEC.2004.1330969

[21] Hafsa Ismail, Kathryn Merrick and Michael Barlow. 9-12 December 2014. Self-motivated learning of achievement and maintenance tasks for non-player characters in computer games. In 2014 IEEE Symposium on Computational Intelligence for Human-like Intelligence (CIHLI). IEEE Explore. Orlando, FL, USA. 978-1-4799-4506-1. DOI: https://doi.org/10.1109/CIHLI.2014.7013386

[22] Simon Demediuk, Marco Tamassia, William L. Raffe, Fabio Zambetta , Xiaodong Li and Florian Mueller. Monte Carlo tree search based algorithms for dynamic difficulty adjustment. In 2017 IEEE Conference on Computational Intelligence and Games (CIG). IEEE Explore. New York, NY, USA. 2325-4289. DOI: https://doi.org/10.1109/CIG.2017.8080415

[23] Joseph Alexander Brown and Daniel Ashlock. Using multiple worlds for multiple agent roles in games. In 2017 IEEE Conference on Computational Intelligence and Games (CIG). IEEE Explore. New York, NY, USA. 2325-4289. DOI: https://doi.org/10.1109/CIG.2017.8080413

[24] Yannick Francillette, Abelkdder Gouaich and Lylia Abrouk. Adaptive gameplay for mobile gaming. In 2017 IEEE Conference on Computational Intelligence and Games (CIG). IEEE Explore. New York, NY, USA. 2325-4289. DOI: https://doi.org/10.1109/CIG.2017.8080419

[25] Cristina Guerrero-Romero , Annie Louis and Diego Perez-Liebana. Beyond playing to win: Diversifying heuristics for GVGAI. In 2017 IEEE Conference on Computational Intelligence and Games (CIG). IEEE Explore. New York, NY, USA. 2325-4289. DOI: https://doi.org/10.1109/CIG.2017.8080424

[26] Luong Huu Phuc, Kanazawa Naoto and Ikeda Kokolo. Learning human-like behaviors using neuroevolution with statistical penalties. In 2017 IEEE Conference on Computational Intelligence and Games (CIG). IEEE Explore. New York, NY, USA. 2325-4289. DOI: https://doi.org/10.1109/CIG.2017.8080437

[27] Andreas Precht Poulsen, Mark Thorhauge, Mikkel Hvilshj Funch and Sebastian Risi. DLNE: A hybridization of deep learning and neuroevolution for visual control. In 2017 IEEE Conference on Computational Intelligence and Games (CIG). IEEE Explore. New York, NY, USA. 2325-4289. DOI: https://doi.org/10.1109/CIG.2017.8080444

[28] Joseph C. Osborn, Adam Summerville and Michael Mateas. Automated game design learning. In In 2017 IEEE Conference on Computational Intelligence and Games (CIG). IEEE Explore. New York, NY, USA. 2325-4289. DOI: https://doi.org/10.1109/CIG.2017.8080442

[29] Junhyuk Oh, Xiaoxiao Guo, Honglak Lee, Richard Lewis and Satinder Singh. Action-Conditional Video Prediction using Deep Networks in Atari Games. In Advances in Neural Information Processing Systems 28 (NIPS 2015).Retrieved April 21, 2019. http://papers.nips.cc/paper/5859-action-conditional-video-prediction-using-deep-networks-in-atari-games