White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/137047/

Version: Accepted Version

Conference or Workshop Item:
Makarovych, Sasha, Canossa, Alessandro, Togelius, Julian et al. (Accepted: 2018) Like a DNA string: Sequence-based Player Profiling in Tom Clancy's The Division. In: Artificial Intelligence and Interactive Digital Entertainment Conference, 13-17 Nov 2018. (In Press)

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Like a DNA string: Sequence-based Player Profiling in *Tom Clancy's The Division*

Alessandro Canossa[a], Sasha Makarovych[a], Julian Togelius[b] and **Anders Drachen[c]**
[a] Ubisoft Massive Entertainment, Drottninggatan 34, 211 41 Malmö, Sweden
[b] Tandon School of Engineering, New York University, 2 MetroTech Center, Brooklyn, NY 11201, USA
[c] Digital Creativity Labs, University of York, The Ron Cooke Hub, Heslington, York, YO10 5GE, United Kingdom

Abstract

In this paper we present an approach for using sequence analysis to model player behavior. This approach is designed to work in game development contexts, integrating production teams and delivering profiles that inform game design. We demonstrate the method via a case study of the game *Tom Clancy's The Division*, which with its 20 million players represents a major current commercial title. The approach presented provides a mixed-methods framework, combining qualitative knowledge elicitation and workshops with large-scale telemetry analysis, using sequence mining and clustering to develop detailed player profiles showing the core gameplay loops of *The Division*'s players.

Introduction

The analysis of player behavior in digital games, often referred to as Game Analytics (El-Nasr, Drachen and Canossa 2013; Bauckhage, Drachen, and Sifa 2015; Yannakakis 2012), has become a cornerstone both for game development and games research, thanks to the introduction of telemetry-based behavioral tracking.

Behavioral telemetry can be applied broadly in development contexts to inform game design, live operations, monetization, balancing, debugging, cheat detection, and not the least permits detailed as well as large-scale investigations of human behavior and psychology, to mention just a few of the application domains. The research area has enjoyed a substantial amount of attention in recent years due to the wide availability of data from games, broad application space and the direct value of behavioral analysis to the industry (El-Nasr, Drachen and Canossa 2013; Nozhnin 2013; Runge et al. 2014; Hadiji et al. 2014).

Game Analytics as a research domain is too young to have well-defined boundaries, similar to associated areas of games research such as Game AI (Yannakakis, G. N. and Togelius, J. 2018) and Games User Research (Drachen, Nacke and Mirza-Babaei 2018). It is also a domain where the pace of innovation is incredibly fast. Furthermore, the vast majority of the research being conducted takes place in the industry, and only limited portions of this is publicly available due to the valuable and confidential nature of insights derived from behavioral analytics. While this is arguably exciting, it also means that situating research within the state-of-the-art of previous work is challenging at best, notably for industry-based research as is presented here.

In this paper, the focus is on the application of sequential behavioral data analysis to build segments, or profiles, of players. Sequential profiling is of interest because as an approach to segmentation, categorization or profiling of players, it integrates a historical viewpoint of the players' actions as they happened in sequence providing more contextual grounding. Snapshot profiling, typically based on aggregate data, is more common but only provides information about the state of the players of a game across the period of time covered by the snapshot (Drachen et al. 2012; Sifa, Drachen, and Bauckhauge 2018). To take an example: one player opens the inventory and tweaks the loadout before a hostile encounter, another player does so only after the encounter. Aggregating the data, the behavior of the two payers is identical, but a key piece of information is left out: the first player displays strategic thinking while the second regroups only after the encounter. Sequential analysis has proven effective at finding patterns in the action space of players (e.g. (Wallner 2015b; Kang, Kim, and Kim 2014; Kastbjerg 2011; Pirker et al. 2016), but as yet there has been no examples of this approach applied in practical game development contexts.

In the work presented here, sequence analysis for the purpose of building behavioral profiles of players towards informing game design is applied to the case of the major commercial game *Tom Clancy's The Division* (Ubisoft, 2016) [in what follows abbreviated to "The Division"], a game with 20 million players. This work presents a framework for how sequence analysis can be performed, as well as how it can be implemented in a commercial development context, expressed through the case of *The Division* and its design team (who noted the sequences are like a player behavior DNA, hence the title of this paper). The framework and method is flexible and generalizable and is designed as a series of steps that can in principle be applied across any game production. The new algorithm that constitutes builds on existing well-known algorithmic elements. Unlike previous academic research on sequence analysis or behavioral profiling in games, the work presented here integrates the developers of *the Division* from start to end, and includes an evaluation

Copyright © 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

of the usefulness of the presented models by the production team. The specific goal is to generate behavioral sequence profiles of players that can be applied directly by the game team to inform design. The work presented thus also provides a step towards addressing one of the key challenges in Game Analytics, namely how to communicate the results of player behavior analysis to the stakeholders who need to convert these results into actionable insights for the game (El-Nasr, Drachen and Canossa 2013).

Background

The work presented here on sequence mining of player behavior as a tool for informing game design falls in the space between previous research strands on **behavioral profiling, behavioral analysis** and **data-driven design insights**, but leads into the space of personalization, recommendation and prediction as well. Along these vectors, a substantial number of papers have been released in the past decade either directly on games or of relevance to behavioral analytical work in games, and due to space restrictions we will here focus on those most directly related to the work presented here.

Focusing on the investigation of **sequences of behavior actions** to construct player profiles, Yang et al. (Yang and Roberts 2014) used a graph sequence approach to extract patterns of tactics used by players in the Multi-player Online Battle Arena (MOBA) game *Dota 2*. The authors describe combat in *Dota 2*, i.e. gameplay, as a sequence of graphs, computing graph metrics as features. The patterns were fed into a decision tree for generating combat patterns. These were then analyzed for how predictive of success they are using frequent subgraph mining (Harrison et al. 2013). Also focusing on esports, Kastbjerg 2011 combined frequent sequence mining with clustering in order to visualize the spatial form of common sequences of player actions in the multiplayer MOBA *Heroes of Newerth*. The work was focused on improving basic heatmaps. It focused on sequences of actions at the detailed level, typically limited to a few seconds, rather than building higher-level models of behaviors. Kang et al. (Kang, Kim, and Kim 2014) similarly focused on repetitive action sequences in games.

Analysis of player behavior as a sequence of behaviors was also conducted by Wallner (Wallner 2015b), who emphasized frequencies of actions as well as the sequential relationship between actions of players, employing Lag Sequential Analysis to determine the significance of sequential transitions in *Heroes of Newerth*. The analysis operates at the level of individual actions and thus the sequences analyzed are short-term, but highlights the application of sequence analysis in exploring patterns and thus strategies used by players. Wallner (Wallner 2015a) also applied the approach to a *StarCraft 2* build orders, discussing the influence of support levels and sequence lengths. Leece & Jhala (Leece and Jhala 2014) similarly applied sequential pattern mining in *Starcraft: Brood War* towards exploring short-term and long-term goals. Finally, Drachen et al. (Drachen et al. 2014) investigated spatio-temporal sequences of behavior in *Dota 2*, indicating differences in team movement as a feature of team skill. The idea of dynamic player modeling has also been approached from an agent-modeling perspective, e.g. by (Missura and Gartner 2009) who emphasized player modeling for difficulty adjustment. (Yannakakis and Togelius 2011) lay out a comprehensive framework for affecting various parts of the game based on models of player experience. Such AI-driven agents can take advantage of historical player data for training purposes (Yannakakis et al. 2013). The focus here is not on AI-driven systems so behavioral analysis for training/informing AIs in games will therefore not be covered further.

In summary, the idea of applying sequence mining to investigate patterns in player behavior remains exploratory and there is a dearth of knowledge on how sequence analysis would operate in a practical development context or how specific approaches scale across games (Wallner 2015b). The work presented here advances on the current state-of-the-art in several ways, notably by: 1) Focusing on long-term play, covering the entire play histories of players, as compared to minute action-behavior sequences; 2) Combining sequence analysis with clustering to build behavioral profiles from the sequence data; 3) Providing an evaluation of the actual usefulness of the sequence analysis and behavioral profiling done from within a major commercial game development company.

The Division: Gameplay

Tom Clancy's The Division was released by Ubisoft in March 2016, developed by Massive Entertainment with assistance from Red Storm Entertainment and Ubisoft Annecy. According to Ubisoft, the game broke industry records for biggest first-week launch of a new franchise with an estimated 330 million USD in sales (Varanini 2016). The game uses the Snowdrop Engine and is available across Windows, PS4 and Xbox one. It passed 20 million players across these platforms in March 2018 (Newhouse 2018). It is a persistent, online, single- or multi-player action role-playing game, set in the near future in New York City following a smallpox pandemic. Players take on the role of Special Agents of the Strategic Homeland Division, referred to as "the Division", and are tasked with helping to rebuild operations in Manhattan, investigate the nature of the outbreak, combat criminal activity, and support civilians. The game contains both single-player, collaborative and player-vs-player elements in an open-world environment which makes it similar to other "sandbox" games such as the Just Cause-series or Elder Scrolls-series in terms of complexity and agency offered to the players, as well as in terms of the dynamic weather systems and other environmental elements.

The gameplay is third-person, with an emphasis on shooter-style combat, cover and exploration. Player characters can carry various weapons, and the game has integral cover mechanics. Players can earn experience points and in-game currencies, using the latter to buy gear and weapons and the former to unlock talents and skills. Gear can be bought or crafted. The game features a storyline of missions that involve various objectives relevant to the players' base of operations. Playing missions gives the player points to access new talents, perks and facilities in their base of operations.

A specific part of the game is the Dark Zone, which is where the Player-vs-Player (PvP) elements of the game are situated, similar to the Crucible of *Destiny* (Sifa et al. 2018). The Dark Zone is separate from the primary campaign and features its own progression systems and access to various items. Dark Zone play can be multi-player or single-player with bots (AI agents). Finally, The Division is a persistent game which has received multiple patches, updates and expansions since its launch in 2016.

Method

Overview and Approach

The quality of player segmentation or profiling is directly dependent on at least four important factors: 1) The data that is used for the segmentation, 2) How that data is processed, 3) How the data is analyzed, and: 4) The degree to which the output of the segmentation analysis is actionable by the stakeholders who are supplied with the result. The latter point is an acknowledged area of weakness in applied Game Analytics (El-Nasr, Drachen and Canossa 2013; Drachen, Nacke and Mirza-Babaei 2018).

In particular, we need *good* features of players and play sessions to use for segmentation. "Good" here means that the features are meaningful in terms of gameplay, that they have sufficient variance over the population, and that they are minimally correlated with each other, so that each feature adds information of its own (Wallner 2015b; Agrawal and Srikant 1995; Pirker et al. 2016; Sifa et al. 2013).

What is *meaningful* in terms of gameplay is of course dependent on the game design. The first step in the model proposed here is therefore to consult with the designers of the game and rely on their knowledge to develop the atomic units of analysis. A common strategy in Game Analytics is to treat player actions and system responses as events, and this nomenclature for the atomic units of analysis is also adopted here as the concept of events in a gameplay context is intuitively understandable across analytics and design/production. While the events of interest can be rare occurrences and thus can be treated on their own, it will generally be the case that events are too frequent and too undifferentiated to tell us much about the players on their own—it is the sequential combination of them that matters in terms of providing patterns of player behavior, such as the sequence: *leaving a supply base - entering a battlefield - entering a firefight* (Pirker et al. 2016; Wallner 2015b; Sifa et al. 2013). The goal is to end up with event sequences that provide insights into a players behavior across the activities in which the player is engaged and how these change as a function of playtime, for example by indicating a specific playstyle (Drachen et al. 2012). We suggest using the number of times a particular subsequences were exhibited by a player as features for that player.

Frequent sequence mining, when applied to behavioral telemetry in games, runs the risk of providing a very large number of sequences for any reasonably large set of player traces (this was exemplified by Kastbjerg 2011). As most unsupervised learning algorithms work better with fewer features, and too many features will make the analysis incomprehensible, we need to decrease the number of features. For this we use information theory (Mackay 2003): We find a small set of sequences which has minimal cross-correlation and where each sequence has maximum entropy through an iterative process. Each of these sequences becomes the basis of a feature, in such a way that the feature value for a player is the normalized number of times the player has exhibited that feature. Thus armed with a dataset of players represented as feature vectors, we can then apply several agglomerative and partition-based clustering methods to find clusters of players with similar behavior. In the current case multiple methods were applied. When applying the framework to a game a choice needs to be made about what aspects of behavior to emphasize, e.g. central tendencies vs. extremal behavior (Sifa, Drachen, and Bauckhauge 2018). The resulting segments then need to be presented to and discussed with the development team in such a way that they can be applied to inform game design.

In summary, we propose a framework for sequence-based player segmentation, or profiling, which broadly encompass the following steps: 1) Defining events with design team, 2) Identifying associated behavioral metrics, 3) Extracting and pre-processing data, 4) Sequence mining, 5) Reducing the number of sequences, integrating the design team, 6) Sequence clustering, 7) Presentation and evaluation with design team leading to implementation of the obtained insights and modifications to the design of the game. In the below we will cover each step, focusing on the application to *The Division*.

Defining behavior events

The end result of a sequence segmentation exercise is a set of profiles of players that include typical behaviors expressed as sequences. For these sequences to be meaningful to a design team, they must be based on behavioral events that are of a type that is informative, and which has the right granularity. For example, completing a mission has a higher granularity than pressing a button on a keyboard. In order to identify the meaningful behavioral events, and the granularity at which the analysis would be carried out, the design team of *The Division* participated in a half-day workshop where the goal and approach of the research was presented. Following, 9 designers (level-, system-, progression- and game designers), covering the different design areas in the game were involved. These worked together with the analysts to list and formally define all the possible activities that players can engage with in the game and all the modifies that can be applied. There was a lot of discussion on what level of abstraction to settle on, meaning very low level activities (for example "looking for cover") although informative on playstyle were abandoned because not revealing preferences and motivations. Additionally, considerable efforts are usually devoted to automatically deriving features with higher level of abstraction from low-level data. While that is a necessary step to infer meaningful features, having access to the design team of the game meant that we could bypass the challenges usually entangled with automatic feature reduction and utilize meaningful features flagged by designers. Ultimately, it was chosen to focus on higher level of

abstraction because it was both more informative of player preferences and easier to capture from a telemetry point of view. Higher level means capturing actions such as "beginning a main mission", or "stumbling into random enemy encounters" or "opening the inventory" or "spending time in a safe house". In applying this workshop exercise in other contexts, the purpose of the analysis needs to be considered when determining the type of event and the granularity of events that are to be analyzed. The result of the workshop was a table of 23 activities, divided into four categories (PvP, PvE, Paid Content and Strategic Planning), with 2 sets of modifiers (solo/group play and four difficulty settings), for a total of 54 behavioral events or activities which it was possible to translate into behavioral metrics. Examples include "main mission", "side mission" and "daily mission", i.e. the players completing either type of these missions. The modifiers are contextual metrics which are important elements of gameplay in *The Division* (full list of activities omitted due to space constraints but available upon request).

Data and pre-processing

Following definition of the behaviors to investigate, and the definition of these behaviors in terms of metrics , data were extracted from the collections servers of *The Division*, covering a period of time from March 2016. 10,000 players were sampled randomly across the total population of players, covering more than 1 million events (of the 52 event types defined), each time-stamped with an anonymized player ID. Data were extracted as JSON objects, which were parsed and converted into a flat comma delimited file. The data is aggregated and exists on a static level, meaning for each character there is a slice of data only at the point in time when the data was pulled. If the character has progressed since (e.g. changed weaponry, made more kills, leveled up), the information is not reflected in the data. Data can be aggregated in the form of *sessions*, with a session being defined as the period from the player started playing the game until stopping, with no breaks in between. I.e. players will have multiple sessions unless they only play the game once. Importantly, if players switch characters (four different characters are possible), this initiates a new session as characters have different abilities etc. and thus gameplay if affected.

Sequential Pattern Mining

Following data extraction, we used Sequential Pattern Mining (SPM) to identify sequences of recurrent behavior in the dataset. For this the open-source SPMF Java-based data mining library was employed, which is engineered to support pattern mining (Fournier-Viger et al. 2014), specifically the CM-SPAM algorithm which is flexible in that it allows defining support, length and gaps (Wallner 2002). Sequential Pattern Mining was employed rather than Frequent Itemset Mining (FIM) in order to preserve the ordering of the transactions in the behavior. The goal is to discover subsequences that appear often in a set of sequences, specifically sequential rules. Note that sequential patterns are defined based on their support, wheres sequential rules are based on support and confidence, and more useful for e.g. recommendation purposes (Fournier-Viger et al. 2014; Agrawal and Srikant 1995; Wallner 2002). Following exploration and iteration, support was set at minimum 0.01, length at min. 2 and gap at 0 (i.e. no gap between activities). These user-defined values impact the outcome of analysis and depending on the specific purpose of the analysis these need to be set accordingly (e.g. focus on shorter vs. longer sequences). This resulted in 826 frequent sub-sequences. Following identification of sub-sequences, the next step was to identify the most information-rich of these. Towards this Shannon Entropy (information entropy) was employed due to its attractive cost functionality via its geometric interpretation (Coifman and Wickerhauser 1992). An entropy score was calculated for each sub-sequence. Using this measure, the 30 highest-scoring sequences were selected. The number of sub-sequences to carry over to the next round is a manual choice, and needs to be made with a consideration for the specific purpose of the profiling exercise. In this case the number 30 was decided on based on a consideration for limiting the dimensionality in the clustering phase. Examples of sequences include "Regroup from Main Solo" which consisted of the events "side mission group" and then "safe area". Cross-correlation was performed to investigate redundancy between sub-sequences, i.e. to identify if any of the sequences split the player population along similar planes, essentially showcasing the same behavioral strings. Each of the 30 sequences receives a normalized score between 0 and 1 that represents the frequency of occurrences of that sequence in all sessions. If two sequences are instantiated by the same players with the same frequency, then they are considered correlated.

Here a number of manual choices need to be made but key heuristics for the final selection of sequences are: a) choice of p-value for accepting correlation, b) the length of strings/sequences (longer are preferred) and c) how meaningful the sequences are for the designers. Towards this, the design team representatives from *The Division* were consulted to verify that the 30 sub-sequences were understandable and given a label. For example since all players start at a safe house, if there are two sequences highly correlated but one starts at the safe house and the other does not, then the second sequence is selected. After the cross-correlation check, the 30 sequences were reduced to 13 sequences; this reduction was carried out in cooperation with the team. This was done in an informal setting, but it is possible to formalize such an evaluation stage, for example in situations where sequence pattern mining is applied to different builds of a game or the number of sub-sequences to be checked is large.

Sequence Cluster Analysis

Following the above steps, we are left with a table with the 13 compounded sequences and the number of times each sequence has been used by each player. This normalized by the number of sessions of the player in question to avoid bias imposed by players having put varying amounts of time into playing the game (generalizing, a player with 100 hours of playtime has a larger chance of activating a specific se-

Table 1: AGNES/Ward cluster loadings, forming the basis for interpretation of the sub-sequence profiles (example descriptions in the text)

Sub-sequence	C1	C2	C3	C4	C5	C6
RK	0	1.74	0.06	0.21	0.21	7.25
SRS	0	0.28	0.05	0.93	1.83	0.34
A	0	0.25	0.06	0.31	2.53	0.53
RG	0	0.85	0.04	0.07	0.14	4.2
RsaG	0	1.12	0.06	0.08	0.12	5.2
STP	0	0.19	0.03	0.33	1.98	0.3
RgaG	0	0.65	0.02	0.07	0.09	3.33
MRS	0	0.08	0.03	1.04	0.43	0.13
RGAS	0	0.15	0.02	0.4	1.07	0.27
RMAG	0	0.8	0.00	0.1	0.06	0.64
MTP	0.00	0.00	0.00	0.00	0.00	0.00
MTPR	0	0.68	0.00	0.1	0.08	0.91
B	1.29	0	0	0	0.01	0

quence than a player with 1 hour of playtime). How to normalize sequence activations is again a manual choice, and the approach will depend on the purpose of the specific analysis. Here session number was used as a proxy for playtime because one of the goals was to find commonly and repeatedly used sub-sequences (patterns of behavior).

The choice of cluster model to apply on behavioral telemetry from games is a topic of discussion across Game Analytics in industry and academia. Comparative work has been presented e.g. by Bauckhauge et al. (Bauckhage, Drachen, and Sifa 2015; Drachen et al. 2012). Different models have varying situations they are suited to. Here three models were selected for testing: k-means, HDBSCAN and AGNES/Ward. To conserve space we do not describe these models in detail here, but description can be found in Aggarwal & Reddy (Aggarwal and Reddy 2013). The models have been applied to player behavior clustering previously (see (Sifa, Drachen, and Bauckhauge 2018)). They form a mixture of variance space search strategies (k-mean via partitioning, HDBSCAN is a density-based method and AGNES/Ward is hierarchical), without extending into extremal/convex hull-seeking models such as archetype analysis. For all analyses, elbow plots were used to determine the number of clusters k. From k=2-8 solutions were investigated for all models, to investigate interpretability of different solutions. As highlighted by Drachen et al. (Drachen et al. 2012), interpretability of player behavior clusters is essential to drive adoption. As for the other steps in the framework proposed here, a choice needs to be made about which models to employ based on the purpose of the analysis. For example, if the goal was to locate extreme behaviors, archetype analysis would be a better choice than the models employed here (Sifa, Drachen, and Bauckhauge 2018).

For k-means, a k=3 solution was selected, with distributions across the clusters being highly uneven, with C1 and C3 containing 10% of the players in the sample, and C2 80%. For HDBSCAN, setting a minimum membership rate of 10%, the model returns a k=2 solution with 30% of the players not residing in any of these clusters considered as noise. C1 contains 50% of the players in the sample, C2 20%. Neither of these solutions resulted in a diverse, balanced cluster solution or profiles that are actionable. AGNES/Ward yielded a k=6 solution with highly imbalanced clusters also but better distribution of the players, and was selected as the result moving forward. The cluster solution distributes the players as follows: C1 = 20%, C2 = 9%, C3=60%, C4=5%, C5=4% and C6=1% (cluster loadings across the 13 sub-sequences in Table 1). Clusters were labeled and a description for each provided, as well as a visual representation of the sequences commonly activated by players in each cluster (Fig. 1)..

Results and Validation with Design Team

In order to validate the quality of the sequence clusters obtained, a second workshop was organized with the same members of the design team of *The Division*. Initially, the clusters were presented as a table with a column for each cluster while the rows contained the features used to derive the clusters. The cells comprised values representing the strength of each feature to define that specific cluster. Designers were asked to provide a label for each cluster based on the insights they could gather from the tables. All the sequence based clusters received distinct and informative labels and were discussed. Key points include the uneven distribution of cluster sizes, not in and of itself a problem, but showed that 60% of the players were not committed to any of the particular sequences (C3). This cluster appears to be the funnel from which players enter the game and, as soon as they accumulate enough playtime, they migrate to other clusters (revealed by a repeating the analysis outlined here on a dataset from the same players for April 2016, i.e. the month immediately following the March sample used here). On the other hand, the remaining clusters are strongly characterized by specific sequences. Cluster 1, the *Lone Street Bandits*, engage in the rogue solo-extraction solo-rogue solo sequence. Cluster 2 and 5, *Lone Side Wanderers* and *Social Side Wanderers*, show preferences for side mission while stumbling onto random hostile encounters, differentiated only by playing solo or in a group. Cluster 4, *Lone Main Planner* moves with surgical precision from main mission to safe areas without stumbling into random hostile encounters. Cluster 6, *Social Farmer 2* (similar to cluster 2), is characterized by moving from random hostile encounters to side mission and occasionally also main mission, but always in a group (it was suggested to merge C2 and C6 due to their relative similarities).

Secondly, the six clusters were presented as network graphs (obtained with Gephi) showing all the sequences of actions instantiated most often by the six clusters identified by the initial segmentation (Fig. 1). The designers discovered that two core gameplay loops are revealed for all the clusters: landmark clear - rogue solo -extraction complete and safe area - inventory operation - loadout operation. Besides the two main loops, the activities listed on the top and right side of the graphs showcase very different priorities for each of the six profiles. For example *Strategizer*, *Social Farmer 1* and *Solo Wanderer* (Fig. 1) display very similar

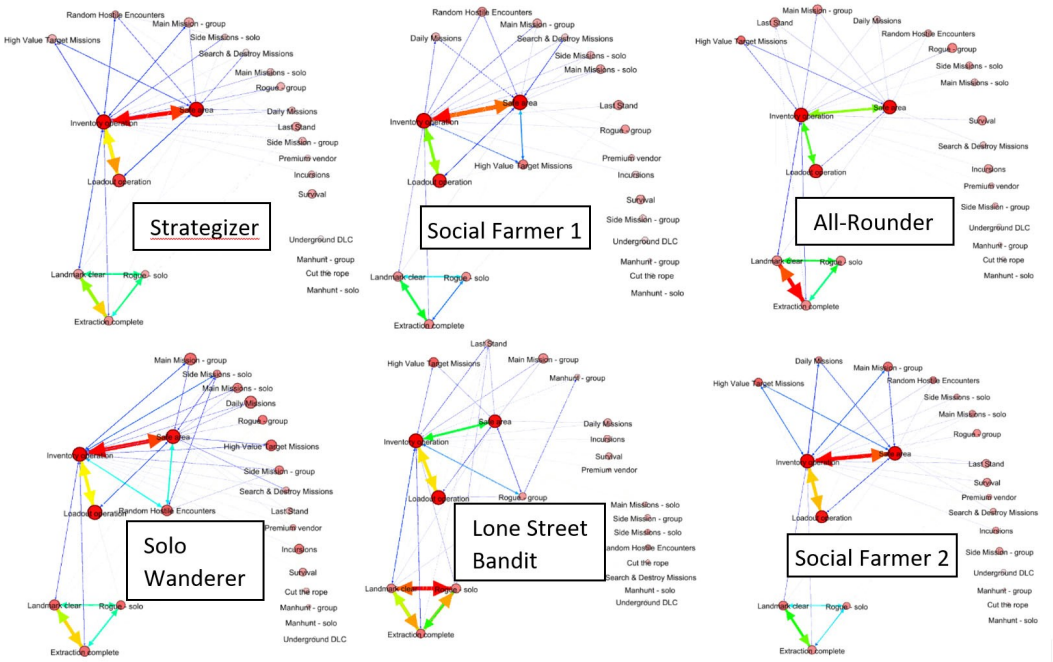

Figure 1: The 6 sequence-based clusters represented as network graphs. Nodes represent the 23 basic activities (w/o difficulty modifiers), edges represent movements between activities. The color and thickness of the arrows indicate how often players in that cluster proceed from one action to the next; size and color of the nodes represents the frequency of that activity being instantiated. All arrows representing movement between actions occuring with a frequency of less than 50% have been removed.

loops, but *Strategizer* tend to follow the loop safe area - inventory operation - loadout operation with high target value missions, which is a mission that displays a high level of intentionality (it does not happen randomly). Meanwhile *Solo Wanderer* follows the loop safe area - inventory operation - loadout operation with random hostile encounters, which is an activity that is triggered randomly while players explore the map. Summarizing, although all six player profiles enact the two core loops, they do so with interesting differences both in terms of the varied intensity of the loops and in terms of which other activities are attached to the main loops. Reflecting on these visualizations, the designers identified 4 themes: 1) The graphs allows them to predict player flows through activities and not just the favorite activities, differently from the aggregated clusters. 2) Monitoring how players access prioritized activities (for example visiting the premium vendor), allows in-depth understanding of funnels. 3) Deeper understanding of funnel allows designers to focus on simplifying access to prioritized activities. 4) These graphs allow comparisons between designers' ideal paths with concrete sequences of actions of players, validating or contradicting the designers intentions. A few of the benefits of activity-based segmentation is that: a) it can be used by designers to analyze gameplay loops, b) it can be used for prediction of desirable and undesirable states (churn, session end, premium vendor visits and premium purchases), and lastly c) it can be used to recommend quickmatches based on preferred activities. Overall the designers agreed that sequence based clusters are both more informative and actionable than traditional aggregate clusters. On a final note, it is important to recognize that the profiles are based on behavioral data and therefore do not inform about intrinsic motivations, personality or any other cognitive states of the players directly. This is a generic limitation on behavioral analytics, but important to keep in mind when discussing how to act on behavioral profiles. This does not lessen the value of sequence-based profiling to game development.

Conclusion and Future Work

Sequence-based profiling or segmentation of players allows us to understand usage or gameplay loops, which provides information about player behavior that aggregate profiling does not (Sifa, Drachen, and Bauckhauge 2018). Sequence-based profiling is in this paper integrated into a flexible framework consisting of multiple phases of analysis, which have then been evaluated in the context of the major commercial title *Tom Clancy's The Division*. Integration of the game's design team provides a means for ensuring the input features and ultimately output profiles are meaningful and can be utilized by the design team. In essence, understanding usage patterns provides information on how to improve games, and the integration of sequences provides the ability to spot the chain of behaviors leading to undesirable events such as churn. Future work will integrate sequence-based profiles in prediction models and build Bayesian networks from network graphs. It is also of interest to explore the dynamic nature/evolution of sequence clusters as a function of time and skill, extending the analysis time frame.

References

Aggarwal, C., and Reddy, C. 2013. *Data Clustering: Algorithms and Applications*. Chapman & Hall/CRC.

Agrawal, R., and Srikant, R. 1995. Mining sequential patterns. In *Proc. of the 11th International Conference on Data Engineering*, 3–14.

Bauckhage, C.; Drachen, A.; and Sifa, R. 2015. Clustering game behavior data. *IEEE Transactions on Computational Intelligence and AI in Games* 7(3):266–278.

Coifman, R. R., and Wickerhauser, M. W. 1992. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory* 38:713–718.

Drachen, A.; Sifa, R.; Bauckhage, C.; and Thurau, C. 2012. Guns, Swords and Data: Clustering of Player Behavior in Computer Games in the Wild. In *Proc. of IEEE CIG*.

Drachen, A.; Yancey, M.; Maquire, J.; Chu, D.; Wang, Y. I.; Mahlman, T.; Schubert, M.; and Klabjan, D. 2014. Skill-Based Differences in Spatio-Temporal Team Behaviour in Defence of The Ancients 2 (DotA 2). In *Proc. of the IEEE Consumer Electronics Society Games, Entertainment, Media Conference*.

Drachen, Nacke and Mirza-Babaei. 2018. *Games User Research*. Oxford University Press.

El-Nasr, Drachen and Canossa. 2013. *Game Analytics: Maximizing the Value of Player Data*. Springer.

Fournier-Viger, P.; Gomariz, A.; Gueniche, T.; Soltani, A.; Wu, C.-W.; and Tseng, V. S. 2014. Spmf: A java open-source pattern mining library. *Journal of Machine Learning Research* 3569–3573.

Hadiji, F.; Sifa, R.; Drachen, A.; Thurau, C.; Kersting, K.; and Bauckhage, C. 2014. Predicting Player Churn in the Wild. In *Proc. of IEEE CIG*.

Harrison; Smith; Ware; Chen; Chen; and Khatri. 2013. Frequent subgraph mining. In *Practical Graph Mining With R*. Oxford University Press.

Kang; Kim; and Kim. 2014. Analyzing repetitive action in game based on sequence pattern matching. *Journal of Real-Time Image Processing* 9:523–530.

Kastbjerg. 2011. *Combining sequence mining and heatmaps to visualize game event flows*. Masters thesis, IT University of Copenhagen.

Leece, A., and Jhala, A. 2014. Sequential pattern mining in Starcraft: Brood War for short and long-term goals. In *Workshop on Adversarial Real-Time Strategy Games at the 2015 AIIDE Conference*. AAAI.

Mackay, D. J. C. 2003. *Information Theory, Inference and Learning Algorithms*. Cambridge University Press.

Missura, O., and Gartner, T. 2009. Player modeling for intelligent difficulty adjustment. In *Proc. of the ECML Workshop From Local Patterns to Global Models*.

Newhouse, A. 2018. The division passes 20 million players across ps4, xbox one, and pc. https://www.gamespot.com/articles/the-division-passes-20-million-players-across-ps4-/1100-6457115/.

Nozhnin, D. 2013. Predicting Churn: When Do Veterans Quit? *Gamasutra*.

Pirker; Griesmayer; Drachen; and Sifa. 2016. How Playstyles Evolve: Progression Analysis and Profiling in Just Cause 2. In *Proceedings of the IEEE International Conference on Entertainment Computing*.

Runge, J.; Gao, P.; Garcin, F.; and Faltings, B. 2014. Churn Prediction for High-value Players in Casual Social Games. In *Proc. of IEEE CIG*.

Sifa, R.; Drachen, A.; Bauckhage, C.; Thurau, C.; and Canossa, A. 2013. Behavior evolution in tomb raider underworld. In *Proc. of IEEE CIG*.

Sifa, R.; Pawlakos, E.; Zhai, K.; Haran, S.; Jha, R.; Klabjan, D.; and Drachen, A. 2018. Controlling the Crucible: A Novel PvP Recommender Systems Framework for Destiny. In *Proc. of ACM ACSW IE*.

Sifa; Drachen; and Bauckhauge. 2018. Profiling in Games: Understanding Behavior from Telemetry. In K., L.; Sukthankar, G.; and Wigand, R. T., eds., *Social Interactions in Virtual Worlds*. Cambridge University Press. 337–374.

Varanini, G. 2016. The division has biggest first week ever for new game franchise. https://news.ubisoft.com/article/the-division-has-biggest-first-week-ever-for-new-game-franchise.

Wallner, G. 2002. Sequential Pattern Mining Using Bitmaps. In *Proceedings of the Eigth ACM SIGKDD International Conference of Knowledge Discovery and Data Mining*, 429–435.

Wallner, G. 2015a. Sequential Analysis of Player Behavior. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, 349–358.

Wallner, G. 2015b. Sequential Analysis of Player Behavior. In *Proc. of ACM CHI Play*.

Yang, P. Harrison, B., and Roberts, D. L. 2014. Identifying patterns in combat that are predictive of success in moba games. In *Proc. of FDG*.

Yannakakis, G. N., and Togelius, J. 2011. Experience-driven procedural content generation. *IEEE Transactions on Affective Computing* 2(3):147–161.

Yannakakis, G. N.; Spronck, P.; Loiacono, D.; and André, E. 2013. Player Modeling. In Lucas, S. M.; Mateas, M.; Preuss, M.; Spronck, P.; and Togelius, J., eds., *Artificial and Computational Intelligence in Games*, volume 6 of *Dagstuhl Follow-Ups*. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. 45–59.

Yannakakis, G. N. and Togelius, J. 2018. *Artificial Intelligence and Games*. Springer.

Yannakakis, G. 2012. Game AI Revisited. In *Proc. of ACM Computing Frontiers Conference*, 285–292.