Behavioural Profiling in Games

A Brief White Paper and Report – Anders Drachen – Digital Creativity Labs

The game industry has access to detailed data about how players are interacting with games. The data can come from a variety of channels are often high-dimensional, time-dependent and potentially very large. *Profiling* users has emerged across multiple data science application areas as a way of managing complex user data, and to discover underlying patterns in the behaviour of the player base. Profiling users allow for a condensation and modelling of a complex behavioural space, and enables action to be taken towards meeting the needs of the users as made apparent via the profiles. Profiles allows us to consider players in a non-abstract, quantifiable way. Building an understanding about who the players are and how they will play, or are playing, the game.

I will in this post outline the background of user profiling and present some notes on a few of the popular types of profiles that can be developed in game contexts. This is by no means exhaustive. Naturally, different types of profiles are used for different kinds of problems, have various strengths and weaknesses, etc.

I am building on work by fellow analysts in industry and academia, so please note that this is not only my work being presented. I include below a reading list with some references. Again, it is not exhaustive, and apologies if I have overlooked important references – please feel free to ping me on twitter (@andersdrachen or @labsofdc) or similar if you want a reference included.

This brief overview report will focus on the conceptual level. Please find a more technical report on user profiling in games [here](). A condensed and updated overview is available as a slide deck [here.]()

Background: Using customer data to inform design

The idea of using customer data to inform marketing and product design has an extensive history in Information Science, where user profiling was developed to deal specifically with the problem of data overload. In games, we easily extract dozens to hundreds of features from direct user-game interaction, and supplement these with data from marketing, attribution, playtesting, social networks and more. To makes things even more challenging, data are usually collected from large numbers of players, from potentially long-term interaction periods, and are typically temporally volatile.

Profiling permits a condensation of the behavioural space so any patterns can be located or hypotheses tested. These are then refined into a format where action can be taken on them and potentially draw inferences as to the root causes of behaviour. In games, there are typically two overall *goals* with player profiling:

1) Correlational: To correlate profiles with specific behaviours such as game completion potential, user experience, monetization, churn, retention, cross-game transportation, cross-promotion, social influence, etc.

2) Inferential: To investigate how and why specific behaviours occur as a function of user traits and/or behaviours.

We can also consider how profiles are *developed*, usually either **bottom-up** or **top-down**. The former is explorative, focused on locating patterns we did not know existed. This approach is useful as soon as we have data (beta, soft-launch etc.), and is usually feature intensive. Top-down profiling focuses on testing hypotheses, e.g. how valid already established profiles are given a new player cohort. This approach is useful post-launch, notably for consistency-testing profiles.

Profiles can be generated either to target **individual** or **groups** of players. Individual profiles seek to discover characteristics of specific people, and is based on data from only that person. Group profiling, which is vastly more common, tries to categorize individuals as a kind of individual – i.e. a type or group. Group profiling is less precise than individual profiling but what we need in practice to manage high-dimensionality datasets. Every group profile will have a *fit* which is the quality of the profile in terms of what it is applied to. Fit is an important component to integrate when considering how to distribute players into profiles, or taking action on players who fall into specific profiles. If a profile is 100% distributive it means that all properties applies fully to everyone in a group, e.g. "all bachelors are unmarried".

In practice, analytically generated group profiles are non-distributive to a greater or lesser degree – and often the latter. This is key concern when considering how to act on profiles, and means that techniques such as soft clustering – which groups people according to their distance to multiple cluster centers (see here for an introduction to clustering techniques) – have value in the daily practice of player profiling. In general, the more detailed we try to make profiles, the less players they apply to, and there is definite element of cost-benefit balancing in play here.

And finally, just to make sure we have the bases covered, player profiles can also be considered based on the information they are built from. Two core types are **protean** and **data-driven profiles**. The latter is based on actual behavioural, attitudinal or other data, while the former is based on theoretical models and design. Data-driven profiles are based on quantitative data, and can be developed from the earliest user testing – they are ideally updated throughout production and during post-launch. Protean profiles are based on theoretical models and commonly used in design. They can be defined from day 1 but importantly must be kept updated to remain useful, which means feeding in design changes and user testing data on a continual basis. They must also be integrated across the team to ensure coherence in their use.

The process of building profiles rests on well-established guidelines for knowledge discovery in IT, irrespective of the specific algorithms or models used for pattern recognition (ranging from simple but effective tools such as cohort analysis to machine learning). In Information Systems, these four basic steps are commonly used:

1) Discovery: A knowledge discovery process is performed to provide sets of correlated data for profiling, i.e. information about which patterns and correlations we see in the data. For example, that kill/death ratio appears to be important to progression in a FPS.

2) Selection: We decide which patterns to use and which behaviours to employ in the further work with developing profiles. For example, if we are interested in churn, we use patterns showing correlations between behaviour and players leaving/staying in the game. Via experimental work we can also investigate causal relationships. Various types of machine learning algorithms can be employed to search the variance space, with clustering being a popular example.

3) Interpretation: In this step we define the profiles. This can be done in a variety of ways, but a sharp eye on the application is important. This is an often over-looked or under-prioritized phase leading to problems in the fourth step.

4) Application: This vital step involves taking action on the information contained in the profiles. This step is possibly the most difficult to execute in practice at it often involves communication between stakeholder groups that speak completely different languages.

The process is of course cyclic. Players change behaviour, the composition of the population changes over time, as does game design in persistent or semi-persistent game. So profiles should be continually updated.
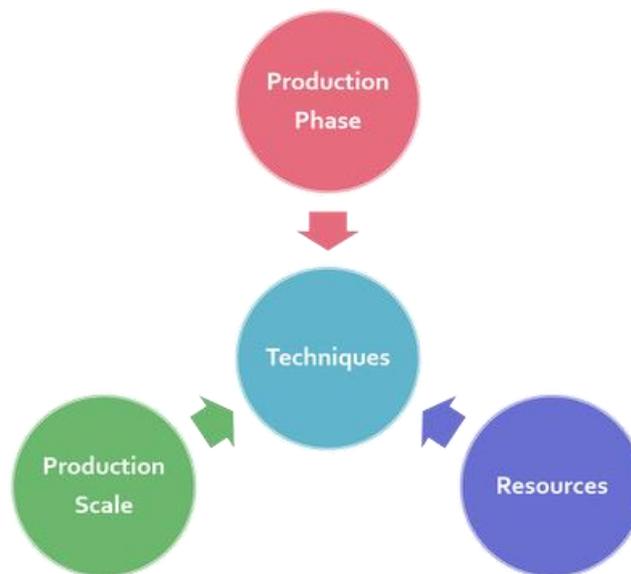
Figure 1: There are numerous techniques available for profiling across qualitative and quantitative approaches. Which technique to use for profiling is determined by a variety of factors, with 3 of the most important being: a) How many resources we have available for profiling; b) The type and scale of the production in question; c) The phase the production is in. Scale + Phase of Production Cycle determines what data is available, and thus what profiling tools we can use.
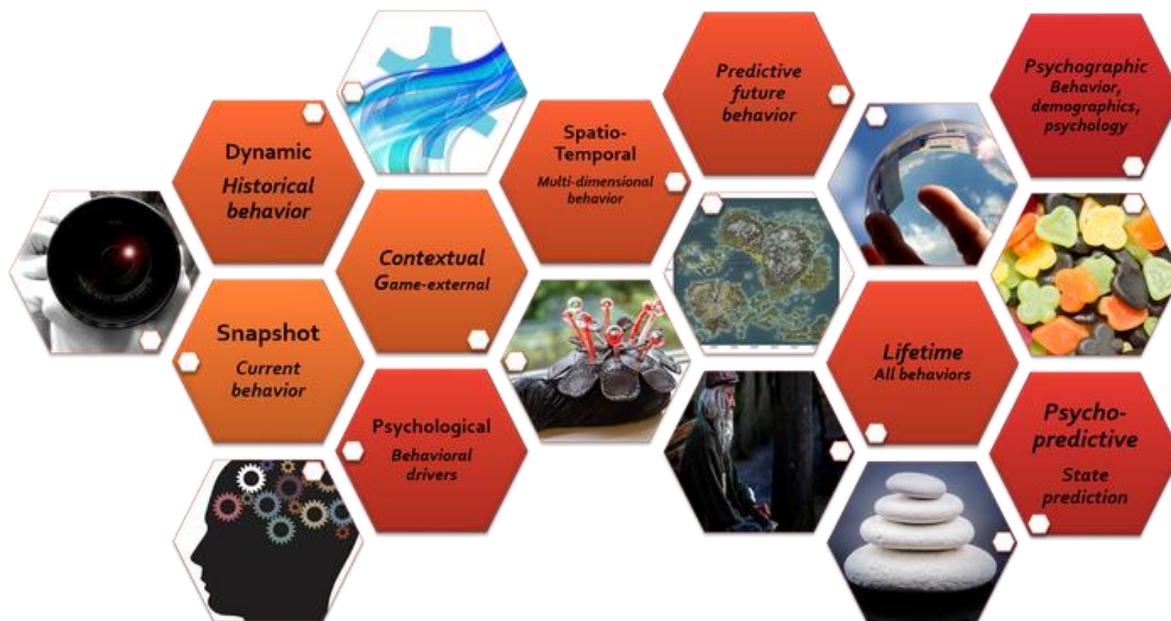
Figure 2: There are different categories of behavioural profiles being utilized in game development. New ideas arrive all the time, and there is a degree of overlap between them, and they are often mixed in practice. There are no hard and fast rules for these, and indeed the terms are used interchangeably across companies and domains.

We are never finished with profiling our players. It is also worth noting that profiling at all levels is not an objective process. We always make choices, the algorithm or model, how data are pre-processed, outcomes interpreted etc. Because of these choices, there is the potential for bias and bad decisions at all of these steps.

Focusing on data-driven profiles, there are a number of different types common in game analytics these days (see illustration).

Snapshot profiles
Snapshot profiling is focused on developing an understanding of the patterns of behaviour as they occur at the operational level. The data used for snapshot profiling are typically aggregate metrics about the players and/or their behaviour. Generally, historical data are not used but rather information about the state of the players at the present. Typical examples include dimensional reduction of high-variety datasets about player characters in Massively Multi-Player Online Games (MMOGs) or other online multi-player games, in order to obtain an understanding about the composition of the current player base.
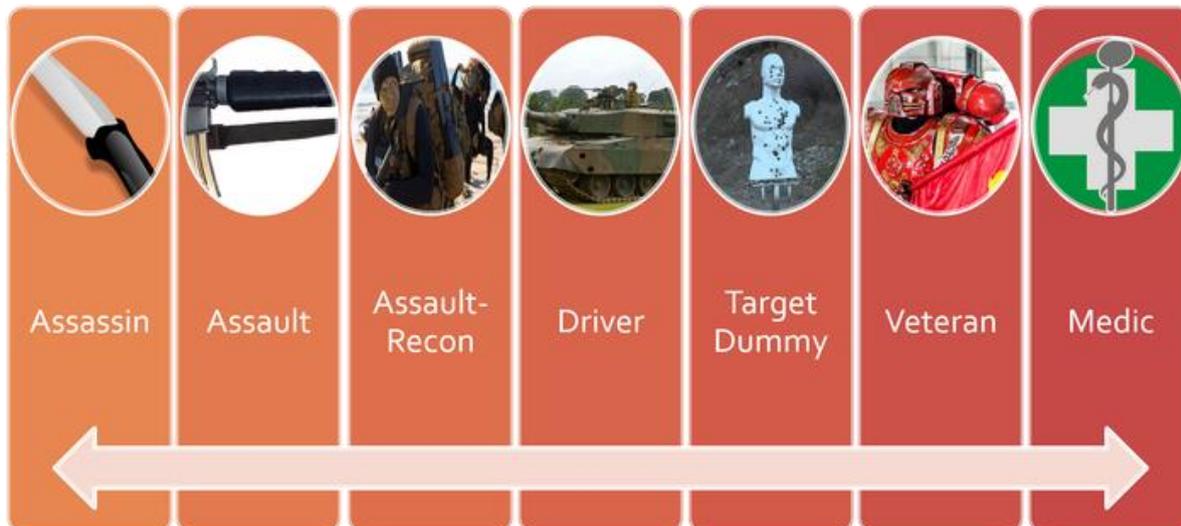
Figure 3: Example of snapshot profiles from Battlefield 2: Bad Company 2, using Simplex Volume Maximization as the clustering model. Analysess revealed seven different profiles in the data, labelled as you see here to add a descriptive element. Each represent different playstyles, and different things that keep them in the game. For example, the "Driver": drives, flies, sails – all the time and favors maps with vehicles. The "Assassin": kills – afar or close – and always walks – favors small maps. For more information see this paper.

Dynamic profiles
Player behaviour change as a function of time. Furthermore, persistent games which sees the same players interacting with a game over potentially long temporal periods, experience a constant change in the population of players. This is notably the case for games which have persistence as a key design factor in order to support F2P business models. Games themselves can also change over time, for example via patches, updates or expansions. These three factors jointly mean that profiles generated based on snapshot data have a limited period during which they are valid as representations of the player base. It is therefore increasingly common to see player profiles being iteratively generated as a function of set time intervals, e.g. 24 hours. While the underlying unsupervised machine learning methods are similar for snapshot and dynamic profiling, the latter are constantly regenerated and additionally permit historical viewpoints on changes in the behaviour of the players (or systems), and also acts as a starting point for predictive analytics. Check out e.g. this paper for an introduction to clustering for profiling and this one for an example in action.
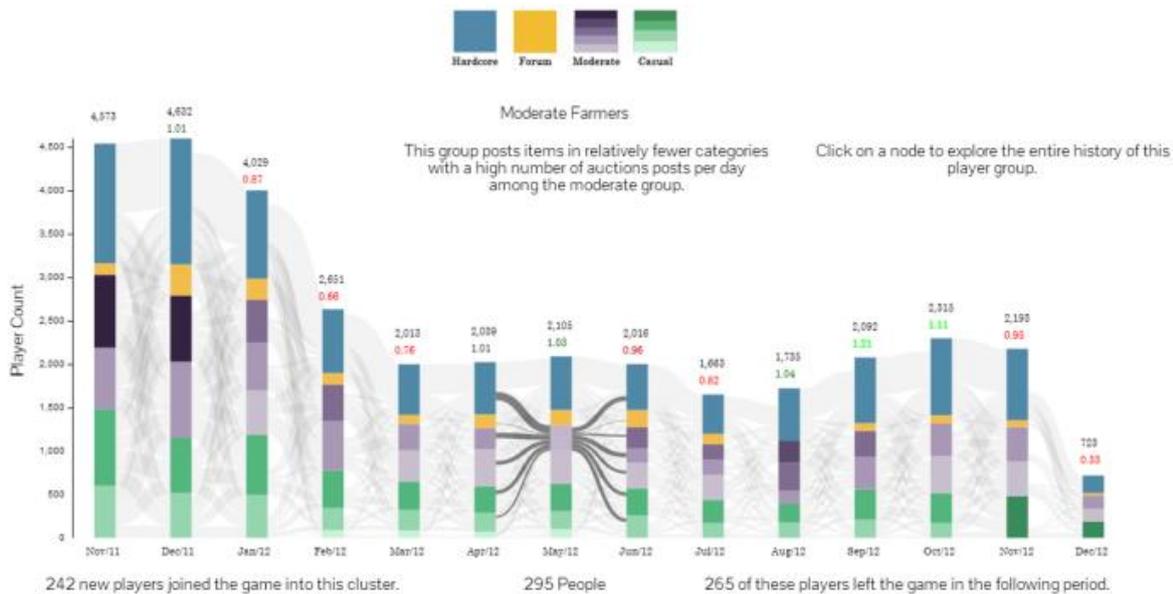
Figure 4: *Example of [dynamic profiling](#) for the MMORPG Glitch. Each column represents a months worth of game data. For each month, cluster analysis is done on selected features relating to the use of the games auction house feature and other economical aspects of the game. A Sankey diagram is used to visualize the flow of players within or between profiles as a function of time. The same approach can be used to investigate [game progression](#) for example.*

Contextual Profiling
Contextual profiles focus on investigating patterns outside the game environment itself. The focus is on understanding the context of play and entry points to play, and there are a variety of potential data sources that can be used including location of play data, attribution data, social media, physical movement, other games/apps installed, demographics, geographic information, etc. The methods being employed also varies substantially, but include segmentation, clustering, attribution models, behavioural funnel analysis, qualitative methods (e.g. social presence impact), etc.

Spatio-temporal Profiling
Spatio-temporal profiling focuses on investigating how people use the virtual environment, notably in 3D games, but can also find uses in evaluating the use of 2D interfaces.

All games contain spatial and temporal elements, whether digital or non-digital. Especially in 3D digital games, the spatial dimension is often important to the perceived experience of the game. Furthermore, spatial navigation and positioning are key gameplay elements in many games. A number of approaches for e.g. trajectory analysis and –classification have been adapted for use in Game AI and in Game Analytics, used e.g. to detect bot programs, study player tactics or to train AI bots. Behavioural analysis can be carried out without considering the temporal and spatial dimensions of play, however, it is often necessary to include one or more of these in order to build the required insights.

Snapshot profiling can be done without historical or spatial data, whereas dynamic profiling often ends up providing temporal patterns. Similarly, predictive modeling requires temporal information. Neither profiling approach requires spatial information, however. Spatial behavioural data are usually only included when needed given the purpose of the analysis. Additionally, spatio-temporal game analytics can be cumbersome and require that interpretation is performed in relation to the actual virtual environment. Ignoring this step in the analysis cycle lead to the risk of misinterpretation of the root causes of the observed behaviours.

The terms "spatial analytics" or "spatio-temporal analytics" in the context of computer games are usually describing the evaluation or analysis of the spatial component of player behaviour, obtained via telemetry sources. For example, the X, Y, and Z coordinates of the location of a player, as well as the time, whenever that player dies, fires a weapon, accepts a quest, punches an opponent, etc. –when an event of any type occurs. In some types of games, it may not make sense to use specific coordinates, but rather information about which zone or area the player is in when a given event occurs. For mobile games, the coordinates of the position of the player in the real world can also be of interest. These are all examples of spatial information that come attached to specific player behaviours.

In practice, spatial analytics often involves a temporal dimension as well, and the two terms "spatial" and "spatio-temporal" are often used to denote the same analytical approach. The focus is on investigating how people use the virtual environment, and capturing the actual dimensions of the user experience. We also use this kind of profiling to investigate if the spatial behaviour of a specific group of players varies. There is a wealth of information to be gathered from GIS, game AI, traffic analysis, Architecture, Ecology and other areas where the movement of entities is important. The methods being employed typically include trail-based clustering and action-sequence mining.
Spatial profiling can of course be adopted at different scales – from a small section of a game level to an entire map.
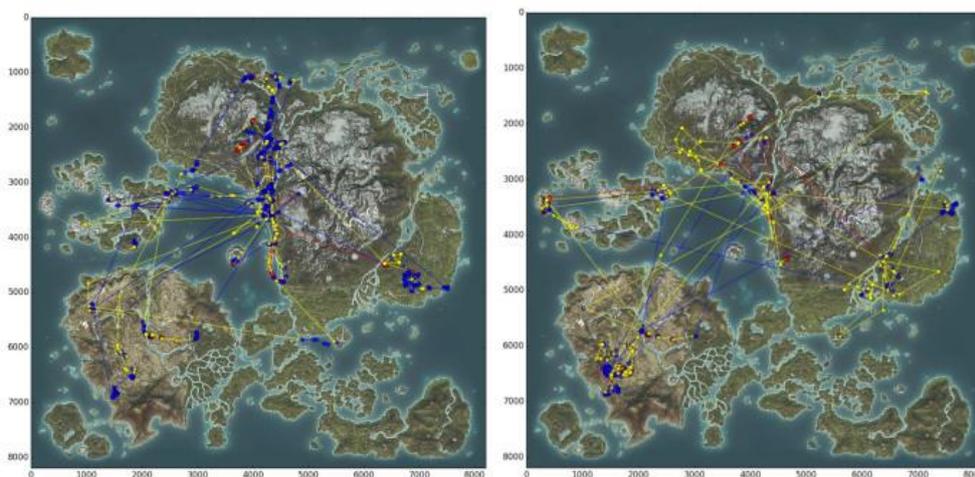
Figure 5: *Examples of spatio-temporal group trails from a 3D game world. Here we use spatio-temporal clustering to develop profiles that integrate both non-spatial and spatial features. It is a great example of where the spatial dimensions were useful. We had two groups of players with*

identical profiles, but when we look at their spatio-temporal behaviour, there are marked differences. The left group mainly experience 3 areas of the game, and use flying vehicles to get around. The right group use the parachute mechanic to get around the map, and focus their time and energy on just one main location. They also engage enemies in different locations, with the left ones engaging in varied locations, the right ones mainly on the largest landmass. Without including the spatial element, we would have been able to catch these differences.

Predictive Profiling

Building predictive models is a key area in business intelligence in general, and predicting player behaviour important in persistent games of any kind. As we all know, F2P games generally see only a small fraction of the players stay engaged with the game for a long time, and a similarly small fraction monetize via IAPs. Therefore, predicting which players that will stay engaged and/or monetize (building implicit profiles), or otherwise provide value, is of key concern in this sector.

Psychological Profiling

Telemetry data are only one source of information about players, and a relatively recently introduced one at that. While the focus here is on telemetry-driven approaches, it should be mentioned that behavioural profiling has an extended history based on information derived from user-testing, surveys, marketing data, etc. The idea of tying in observations from gameplay to profiling, or use gameplay behaviour as the sole basis for profiling, was introduced much later. Models focusing on player motivations, personality, etc. have been around for over a decade, but only recently have we started correlating these with telemetry data. Similarly, the idea of using telemetry data to draw inferences about player psychology is also relatively recent, and there is limited publicly and/or systematic knowledge available at the time of writing.

The value in building psychological profiles of players is clear: it allows us to understand why people play as they do. With behavioural telemetry we can only draw inferences. Psychological profiling involves two steps: 1) model building using experimental methods and psychological models; 2) testing of these models on large scale behavioural data. So we combine user research techniques (observation, interviews …), operationalized psychological theories and behavioural telemetry (or contextual data).

Lifetime Profiling

Lifetime profiling focuses on building an understanding of the player throughout their lifetime, i.e. the period during which they interact with one game or a portfolio of games. There are two basic goals, either anticipating the sum total of a feature associated with a player, for example customer life time value (CLTV); or predicting behaviour across many aspects, multiple future touchpoints e.g. purchases, likely obstacles to progression, predicted social activity, engagement etc. In games, we have a good understanding of lifetime profiling when it comes to purchase prediction and social prediction, although these are far from solved problems, but less strong understanding outside these areas. The methods being used here are mainly supervised machine learning techniques good for sparse datasets such as various supervised models, recurrent nets and tensor flow models.

Psycho-predictive Profiling

This is a very new idea in games contexts, essentially focusing on predicting the future psychological state of the player. There are no publicly available examples at the time of writing this, but the core idea is to build an understanding of the psychological state of the player at points in the future. The hypothesis is that if we can understand what e.g. the motivations for play are at a given point now, we can build an understanding of motivations for similar instances in the future. The value of these kinds of models is obvious: understanding the psychological state of the player provides information for revenue optimization (item sales, ads) and game adaption (e.g. stress prediction and prevention).

Summarizing and takeaways

Summarizing, player profiling is an incredibly useful tool as it allows us to bridge the gap between the users and analytics. Profiles can provide a deep understanding of the players, and serve as the basis for a range of analytical techniques including prediction. There are a treasure trove of techniques available which spans the range from descriptive methods to advanced machine learning, and this means that profiling as an exercise to glean value from player data is open to everyone. In practice, we can get very far with simple methods, even in complex situations. Profiling requires control of every step of the process however, and well-considered application of the profiles in practice.

Summarizing, here are the condensed takeaways from this rather lengthy post:

1. **Keep the focus on the users as people:** needs, requirements, user experience, behaviour, psychology … -> **UX is King**

2. Integrate a deep feedback loop between **analytics**, **user research, design** and **strategy**

3. **Need-driven:** cost resources to build profiles, must be justified by a need

4. **Verified** by or driven by data (behavioural, attitudinal, model …)

5. **Dynamic** and time-sensitive (*timed for their purpose*)

6. Easy to **explain** and to take **action** on by the relevant stakeholder

I also have a couple of warnings:

1. Profiling is not an **objective** process. There are choices being made: algorithm, normalization, pre-processing, interpretation … which means there is a *potential for bias and bad decisions at all steps*

2. Profiling can **integrate varied data** sources – even theory – take care cross-correlating

3. In practice: **algo-/model-generated** group profiles are non-distributive

4. Key challenge is (always…) **feature selection**

5. **The lure of machine learning**: descriptive stats and simple profiles: often fast & surprisingly useful.