Developing Prototype Player Personas from Game Preferences

Joni Salminen
Qatar Computing Research Institute; and Turku School of Economics
Doha, Qatar
jsalminen@hbku.edu.qa

Jukka Vahlo
Turku School of Economics
Turku, Finland
julavi@utu.fi

Aki Koponen
Turku School of Economics
Turku, Finland
aki.koponen@utu.fi

Soon-gyo Jung
Qatar Computing Research Institute
Doha, Qatar
sjung@hbku.edu.qa

Bernard J. Jansen
Qatar Computing Research Institute
Doha, Qatar
jjansen@acm.org

ABSTRACT

We generate player personas from game preference survey data using the system and methodology of automatic persona generation (APG). The purpose is to demonstrate the potential of data-driven technologies for segmenting players by their game preferences. The resulting prototype personas are particularly intended for game marketing purposes, e.g. targeting gamers with social media advertising. The personas can also be enhanced by additional data to provide deeper insights.

CCS CONCEPTS

• CCS → Social and professional topics → User characteristics

KEYWORDS

Player Personas; Automatic Persona Generation; Gaming Preferences

ACM Reference format:

FirstName Surname, FirstName Surname and FirstName Surname. 2018. Insert Your Title Here: Insert Subtitle Here. In *Proceedings of ACM Woodstock conference (WOODSTOCK'18). ACM, New York, NY, USA, 2 pages.* https://doi.org/10.1145/1234567890

1 INTRODUCTION

Although personas are shown to be useful to many use cases in design, software development, and gaming [5,16], their application is prohibited by both lengthy time and the high cost of creation. Therefore, many startups and small companies cannot afford the investment of a full-scale persona project. Consequently, these organizations are generally not able to leverage personas.

Automatic persona generation (APG) is a methodology that brings personas to the reach of more people [1,2], enhancing customer-driven decision making and helping organizations achieve their goals in competitive markets. APG is defined as a methodology and system for creating personas from digital data [1,2]. It is specifically developed to address the limitations of manual persona creation. Personas from APG are (1) rapidly created, (2) representative, as they are based on large amounts of quantitative data, and (3) can be easily periodically updated through re-processing fresh datasets obtained via analytics APIs.

In this work, our objective is to apply the APG approach to a new context, documenting the first case of applying automatic persona generation to develop player personas. With this research, we also address the shortcoming of previous player persona research, namely the fact that the developed personas in previous research [8,10] are missing personified attributes (e.g., picture, name and other human attributes). We collect survey data on players' favorite games (i.e., game preferences) and apply the APG methodology to develop data-driven personas from this dataset.

2 RELATED WORK

2.1 Player Personas

Player personas are intended for stakeholders interested in gaming audiences, including game designers and developers, marketers, and players who wish to understand other players. Particularly in the gaming context, personas can be used to represent player archetypes involving player attributes and game choice preferences [16,17].

In previous research, players personas have mostly been created by using game analytics tools, which typically focus on the effectiveness of level design, player behavior, monetization, and game performance at large. Many companies combine the former with play testing, questionnaires, interviews, and psychometrics that are utilized across game development phases.

WOODSTOCK'18, June, 2018, El Paso, Texas USA

https://doi.org/10.1145/1234567890

For example, tuning the difficulty level of a game is known to be an important factor for successful game design, and companies use telemetric data and game log data to find an appropriate level of difficulty to fit the needs of their core audiences [6].

Drachen et al. [7] have applied player persona approach to identify player profiles based on player behavior factors, e.g., completion time and number of deaths. The challenge for game behavior-based player personas is that it is difficult to extend the identified constructs to other games [6]. This challenge is largely because procedural player personas are typically constructed based on specific characteristics of a single game [8,10] providing limited tools for understanding sustaining game preferences and game choice of particular player segments. In order to generate player personas for predicting game choice, the analyzed data should be based on players' game preferences instead of their behaviors when playing a single game.

2.1 Data-Driven Personas

Data-driven personas are based on empirical data of large quantities, often focusing on behavioral metrics and demographic attributes [2]. In particular, automatic persona generation is a data-driven persona creation methodology that addresses the limitations of qualitative persona creation reported in the literature [4]. The benefits of APG are shown in Table 1.

Table 1: Benefits of automatic persona generation

Manual personas	*Automatic personas*
Low sample size ("small data")	High sample size ("big data")
Qualitative data	Quantitative data
Slow to create (typically taking months)	Fast to create (typically taking days)
Unresponsive to changes in user preferences	Responsive to changes in user preferences
Expensive	Affordable

Data-driven personas can be created from system logs and organizational records describing the users or customers [3]. For example, Molenaar [12] grouped 400,000 clickstreams into common workflows and classifying users into the workflows.

Using a quantitative approach, Zhang et al. [18] analyzed clickstream data with hierarchical clustering to identify common click flows and generate data-driven "personas" with names. However, their personas only include a manually given name but no other personified information, such as age, gender, location, and motives. This limitation also applies to procedural player personas that are generated e.g. with evolutionary algorithms and neural networks [8]. These personas are manually given names based on their behaviors of playing the game (e.g., "monster killers"). Rather than human beings with name and personified information, these personas are more like virtual agents that capture the behavioral variation within the game playing context [15]. Therefore, creating player personas with personified attributes is an open research question.

3 METHODOLOGY

3.1 Data Collection

We collected data using an online form, in which the respondents were able to choose their favorite games. In addition, the respondents indicated their age, gender, and their preferences in gameplay elements of video games. The respondents were recruited from Finland in 2017 by marketing the online form as a short player type test.

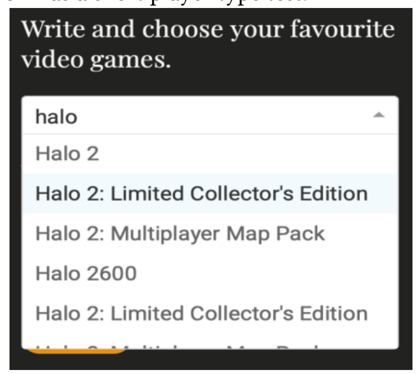


Figure 1: Game preference selection. The field asks respondents to type their favorite video games.

Figure 1 illustrates the form: respondents type into the form field which triggers a query to the underlying database of 130,495 game titles, and presents the respondents autocomplete suggestions. The game titles were obtained from Internet Game Database (IGDB) using the API service (https://www.igdb.com/).

By taking the test on www.kinrate.com, respondents would receive immediately their player type and game recommendations based on similar player profiles to their own. The final dataset includes 195,158 game choices by the respondents (see Table 2), collected via the online form.

Table 2: Description of the dataset

Total number of preferences	*Number of unique respondents*	*Average number of preferences per respondent*
195,158	15,402	12.67

3.2 Player Persona Generation

To generate player personas from the survey data, the APG methodology undertakes several steps:

1. Create an interaction matrix with games as columns and players as rows
2. Apply NMF to interaction matrix to discern latent game preference patterns
3. Choose the representative demographic attributes for each pattern by using NMF weights
4. Create the personas by enriching the representative demographic groups with extra information.

First, we create an interaction matrix where rows are players and columns are games chosen by each player. For example, if Respondent 01 chose "Halo" as their favorite game but

Respondent 02 did not, the matrix would have the value of "1" for the cell <R01-Halo> and "0" for <R02-Halo>.

Because we have p=15,402 players and g=130,495 possible games, this interaction matrix has p x g = 2,009,883,990 cells.

Second, after obtaining the interaction matrix, we group the games by *age* x *gender* combinations. To reduce data dimensionality, we do this grouping by two genders (male, female) and seven age groups: 13-17, 18-24, 25-34, 35-44, 45-54, 55-64, and 65+. These age groups are compatible with the typical age groups available in online analytics platforms [2]. Figure 2 illustrates the process of transforming individual game preferences into grouped instances.

Country	Gender	Age	Game
FI	male	19	Final Fantasy VII
FI	female	52	Halo
FI	male	36	Pokémon Go
FI	male	24	Final Fantasy VII

Country	Gender	Age	Game	Instances
FI	male	18-24	Final Fantasy VII	2
FI	female	45-54	Halo	1
FI	male	35-44	Pokémon Go	1

Figure 2: Data preparation step in automatic persona generation. Individualized data is grouped by age and gender in order to reduce data sparsity.

After obtaining a grouped interaction matrix, we apply non-negative matrix factorization (NMF) [9] for identifying latent game preference patterns. NMF is particularly intended for reducing the dimensionality of large datasets by discerning latent factors. It is widely applied for various tasks, including recommendation systems and feature engineering for machine learning [11]. Figure 3 illustrates the matrix decomposition process of NMF; the resulting patterns inferred from the matrix discriminate the players into different groups based on the variation of their game preferences (i.e., the games they chose).

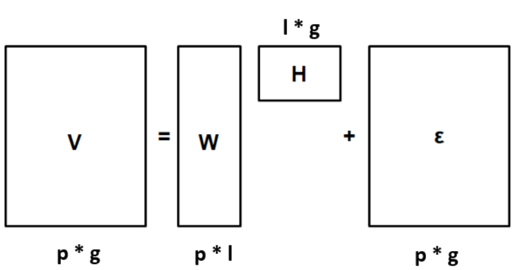


Figure 3: Matrix decomposition carried out using NMF. Matrix V is decomposed into W and H. p denotes players (respondents in the dataset), g denotes games, and l is the number of latent game preference patterns. For this research, we set l = 10, as a relatively low number of personas is a recommended practice [5].

The NMF patterns, therefore, form the basis of the created persona set. Note that because the NMF relies on game preference data that is grouped by demographic attributes, the resulting patterns account both for demographic and behavioral variation in the underlying dataset [1].

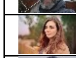
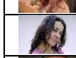
Fourth, after conducting NMF, we choose the representative demographic group for each latent pattern discerned by NMF. This is done according to factorial weights; we simply choose the demographic group with the highest NMF weight as the representative demographic group of the corresponding persona.

Finally, the representative demographic groups are enriched with additional information, including picture, name, favorite games, job, relationship status, education status, and audience size. The photo and name are chosen from the proprietary database of age-, gender-, country-specific names and portrait photos [2] The following information is retrieved from Facebook's audience data using Facebook Marketing API[1]: (1) job, (2) relationship status, (3) education status, and (4) audience size. The values shown for each persona are the most typical values of the Facebook segment corresponding to that persona.

4 RESULTS

Table 3 summarizes all the 10 personas generated from the game preference data.

Table 3: Generated personas. The dataset had only Finnish respondents, so all the personas are Finnish as well.

	Name	*Age*	*Gender*	*Location*
	Joonas	32	Male	Finland
	Jani	44	Male	Finland
	Veera	26	Female	Finland
	Pekka	18	Male	Finland
	Johanna	22	Female	Finland
	Jussi	48	Male	Finland
	Henna	40	Female	Finland
	Marianna	45	Female	Finland
	Jari	57	Male	Finland
	Teemu	15	Male	Finland

Note that the names and pictures of the personas are automatically selected from an internal system database, with tens of thousands of meta-tagged names and thousands of meta-tagged pictures purchased from stock photo companies or commercial photographers.

End users of the APG can use the interface to select any of the generated personas. This shows the full persona profile. An example of the automatically created player persona profile is shown in Figure 4.

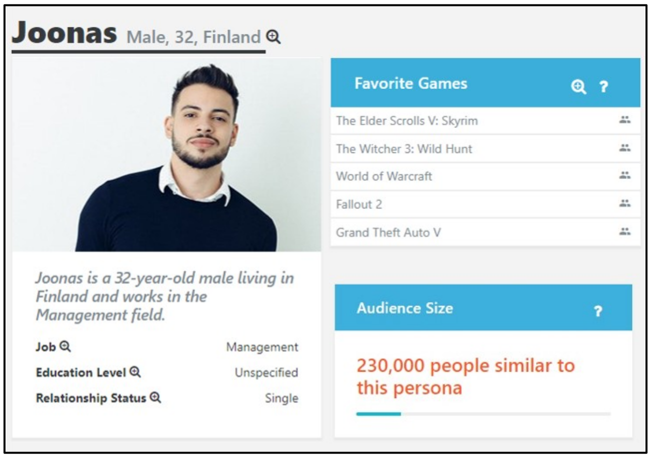


Figure 4: "Joonas" – An example of automatically generated persona profile

To provide algorithmic transparency, the NMF weights for the 10 highest-ranking demographic groups are also shown to end users of the APG system (see Figure 5).

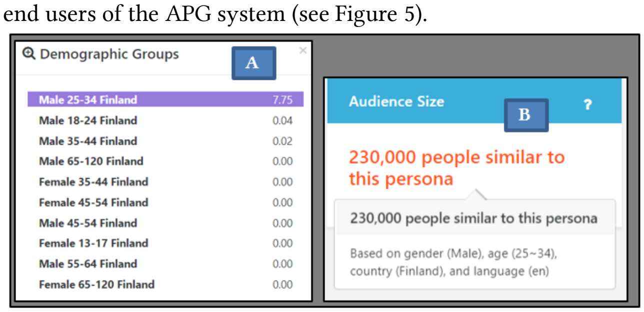


Figure 5: [A]: Top demographic groups of Joonas. The group with the highest weight (numerical value) is chosen as the representative group (unless it has been chosen previously, in which case the second highest is chosen, etc.). [B]: Audience size is retrieved using Facebook Marketing API and targeting criteria corresponding to persona information.

While we leave the formal evaluation of player personas for future work, Figure 6 shows that the generated player personas correspond well with the gender distribution in the raw survey data. Also, note that the APG methodology has been previously evaluated and validated in other contexts, i.e., users' online content preferences [1,2,14].

As far as algorithmic bias goes, research has shown that APG personas accurately inherit the biases present in the data (e.g., male/female imbalance) [13]. However, there is no evidence on the APG imposing a non-obvious bias. Rather, previous research supports the notion of APG as a robust data-driven persona generation methodology that provides personas that accurately represent the source data [1,2].

[1] https://developers.facebook.com/docs/marketing-apis/

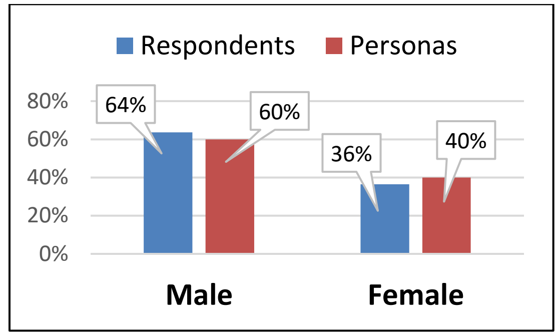


Figure 6: APG generated four female personas out of ten (40%), corresponding well to the gender distribution of the survey data (36%).

5 DISCUSSION AND FUTURE WORK

We call the outcomes prototype player personas because they are lacking information that is typically shown in persona profiles, such as interests, pain points, and quotes [5]. Even though complete, rounded personas should be the end goal, we highlight that even at the current state the system is a step forward in creating data-driven player personas. This is because previous attempts have produced even less empathetic persona profiles, with missing demographic and other personified information. For example, Holmgard et al. [8] produce personas from game-playing behaviors but with no name, face, or personified information.

Ultimately, the type of personas created is dictated by the use case at hand – the creation of automatic personas would particularly suit *game marketing* and *game marketers*, because they provide additional information that can be used e.g. for ad targeting. To provide useful representations for game developers and designers, future research should enrich the prototype personas with extra information. In particular, we plan to use the results from a factor analysis conducted in a previous study [14] to enhance prototype personas with game dynamics, e.g., assault, manage, journey, coordinate, and care. This corresponds to the notion of hybrid personas that combine qualitative and quantitative aspects for in-depth user insights [13].

REFERENCES

1. Jisun An, Haewoon Kwak, Soon-gyo Jung, Joni Salminen, and Bernard J. Jansen. 2018. Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data. *Social Network Analysis and Mining* 8, 1. https://doi.org/10.1007/s13278-018-0531-0
2. Jisun An, Haewoon Kwak, Joni Salminen, Soon-gyo Jung, and Bernard J. Jansen. 2018. Imaginary People Representing Real Numbers: Generating Personas from Online Social Media Data. *ACM Transactions on the Web (TWEB)* 12, 3.
3. Jonalan Brickey, Steven Walczak, and Tony Burgess. 2012. Comparing semi-automated clustering methods for persona development. *IEEE Transactions on Software Engineering* 38, 3: 537–546.
4. Christopher N. Chapman and Russell P. Milham. 2006. The Personas' New Clothes: Methodological and Practical Arguments against a Popular Method. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 5: 634–636. https://doi.org/10.1177/154193120605000503

5. Alan Cooper. 1999. *The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity.* Sams - Pearson Education, Indianapolis, IN.
6. Benjamin Cowley and Darryl Charles. 2016. Behavlets: a method for practical player modelling using psychology-based player traits and domain specific features. *User Modeling and User-Adapted Interaction* 26, 2–3: 257–306.
7. Anders Drachen, Alessandro Canossa, and Georgios N. Yannakakis. 2009. Player modeling using self-organization in Tomb Raider: Underworld. In *Computational Intelligence and Games, 2009. CIG 2009. IEEE Symposium on*, 1–8.
8. C. Holmgard, M. C. Green, A. Liapis, and J. Togelius. 2018. Automated Playtesting with Procedural Personas with Evolved Heuristics. *IEEE Transactions on Games* PP, 99: 1–1. https://doi.org/10.1109/TG.2018.2808198
9. Daniel D. Lee and H. Sebastian Seung. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 6755: 788–791. https://doi.org/10.1038/44565
10. Antonios Liapis, Christoffer Holmgård, Georgios N. Yannakakis, and Julian Togelius. 2015. Procedural Personas as Critics for Dungeon Generation. In *Applications of Evolutionary Computation*, Antonio M. Mora and Giovanni Squillero (eds.). Springer International Publishing, Cham, 331–343. https://doi.org/10.1007/978-3-319-16549-3_27
11. Xin Luo, Mengchu Zhou, Yunni Xia, and Qingsheng Zhu. 2014. An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems. *IEEE Transactions on Industrial Informatics* 10, 2: 1273–1284.
12. L. Molenaar. 2017. Data-driven personas: Generating consumer insights with the use of clustering analysis from big data. TU Delft, Netherlands. Retrieved March 17, 2018 from http://resolver.tudelft.nl/uuid:12d7f261-20b4-4656-93d7-fed2b437aefb
13. Joni Salminen, Soon-Gyo Jung, and Bernard J. Jansen. 2019. Detecting Demographic Bias in Automatically Generated Personas. In *CHI'19 Extended Abstracts: CHI Conference on Human Factors in Computing Systems Extended Abstracts Proceedings*.
14. Joni Salminen, Sercan Şengün, Haewoon Kwak, Bernard J. Jansen, Jisun An, Soon-gyo Jung, Sarah Vieweg, and Fox Harrell. 2018. From 2,772 segments to five personas: Summarizing a diverse online audience by generating culturally adapted personas. *First Monday* 23, 6. Retrieved June 3, 2018 from http://firstmonday.org/ojs/index.php/fm/article/view/8415
15. Jukka Vahlo, Johanna K Kaakinen, Suvi K. Holm, and Aki Koponen. 2017. Digital Game Dynamics Preferences and Player Types. *J. Comp.-Med. Commun.* 22, 2: 88–103. https://doi.org/10.1111/jcc4.12181
16. Jukka Vahlo and Aki Koponen. 2018. Player Personas and Game Choice. In *Encyclopedia of Computer Graphics and Games*, Newton Lee (ed.). Springer International Publishing, Cham, 1–6. https://doi.org/10.1007/978-3-319-08234-9_149-1
17. Jukka Vahlo, Jouni Smed, and Aki Koponen. 2018. Validating gameplay activity inventory (GAIN) for modeling player profiles. *User Modeling and User-Adapted Interaction* 28, 4: 425–453. https://doi.org/10.1007/s11257-018-9212-y
18. Xiang Zhang, Hans-Frederick Brown, and Anil Shankar. 2016. Data-driven Personas: Constructing Archetypal Users with Clickstreams and User Telemetry. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 5350–5359.