Discovering Social and Aesthetic Categories of Avatars: A Bottom-Up Artificial Intelligence Approach Using Image Clustering

Chong-U Lim[1], Antonios Liapis[2], and D. Fox Harrell[3]
[1]Massachusetts Institute of Technology, Cambridge, Massachusetts, USA,
culim@mit.edu
[2]Institute of Digital Games, University of Malta, Msida, Malta,
antonios.liapis@um.edu.mt
[3]Massachusetts Institute of Technology, Cambridge, Massachusetts, USA,
fox.harrell@mit.edu

ABSTRACT

Videogame avatars are more than visual artifacts—they express cultural norms and expectations from both the real world and the fictional world. In this paper, we describe how artificial intelligence clustering can automatically discover distinct characteristics of players' avatars without prior knowledge of a system's underlying data structures. Using only avatar images collected from a study with 191 players, we applied two clustering techniques—namely non-negative matrix factorization and archetypal analysis—that automatically revealed and detected (1) an avatar's gender, (2) regions that appeared to isolate shapes of items and accessories, and (3) aesthetic preferences for particular colors (e.g., bright or muted) and shapes for different body parts. These clusters correlated with players' preferences for character abilities, e.g., male avatars in dark clothes correlated with having high physical but low magic-casting attributes. These findings show that a bottom-up analysis of images can reveal explicit categories like gender, but also implicit categories like preferences of players. We believe that such computational approaches can enable developers to (1) better understand players' desires and needs, (2) quantitatively view how systems may be limited in supporting players, and (3) find actionable solutions for these limitations.

Keywords
avatars, non-negative matrix factorization, archetypal analysis, unsupervised learning

INTRODUCTION

Video games and virtual worlds present the opportunity for players to experience novel settings and environments. Often times, the human takes the role of player characters, or avatars, to exert influence and control within these environments. Players often spend considerable time customizing their avatars (Yee 2006). Some prefer to create avatars that look like themselves (or something/someone they admire) (Boellstorff 2008); others seek to communicate abilities and status (Lim and Harrell 2014; Gee 2003). In many cases, avatar representations are important for both enjoyment and (in-game) practicality. Avatar customization systems often provide a large variety of options for players to construct satisfying representations. Players can clothe themselves in medieval attire, sport different hair colors and styles, or even take on a different gender identity to their own.

Proceedings of 1st International Joint Conference of DiGRA and FDG

However, the technical data structures implementing these avatars may impose infrastructural limitations that fail to support the nuances of real-world identities (Harrell 2009). For example, a player identifying as mixed-race (of multiple heritages) might be limited to picking just one race for an avatar simply because of the lack of support for multiple races at the implementation level. Besides technical limitations, some issues are a result of design choices. Consider players intending to play as characters whose race matches their own. They might encounter tension if those characters were attributed undesirable characteristics by design—forcing them to switch characters to avoid being disadvantaged. These examples highlight the challenge and importance of developing ways to identify such situations from both a developer's and player's perspective. Our approaches seek to demonstrate how such implicit player preferences and social identity constructs—such as racial attitudes and stereotypes that manifest in the virtual world (Ash 2015)—can be revealed.

Motivation: Analyzing Avatars for Player Modeling

Avatars can reveal salient characteristics of one's preferences for performing identity such as gender and other aspects such as behaviors and motivations for play (Lin and Wang 2014). We argue that constructing computational models of how players choose to represent themselves with avatars can provide better models of players. This will enable developers to (1) better understand what their players' desires and needs are, (2) identify how their games may be limited in supporting these players, and (3) have quantifiable and actionable ways to improve their game's designs. We propose using data-driven approaches for constructing such models. This entails collecting and analyzing player avatar data to discover emergent patterns in a bottom-up manner—rather than adopting predefined assumptions—as a way to develop a better understanding of one's players. The advantages of this approach are two-fold. First, it requires no knowledge of the underlying representation used to implement the avatars. This better reflects how players visually perceive in-game characters—they are often not privy to a game's code and instead make inferences based on what they can see. Second, it can reveal novel or unforeseen characteristics of how players choose to use their avatars for identity construction. Such emergent patterns are better indicators of players' needs—since they are not based on designer's assumptions—and can help focus efforts to improve the design of games, customization systems, and avatars alike.

Outline of this Paper's Contributions

In this paper, we present our approach of computationally revealing how different players represent themselves virtually using avatars. While other approaches for eliciting such information exist—e.g., self-reported user surveys—our approach uses artificial intelligence (AI) to automatically discover emergent patterns. This data-driven approach overcomes the limitations of other approaches like survey bias (Bauckhage et al. 2014). Informed by theories of cognitive categorization (Rosch 1999; Lakoff 1990; Bowker and Star 1999) and the psychology of color perception (Berk et al. 1982), we use AI to analyze images of player-created avatars from a custom-built avatar creator (Lim and Harrell 2015c). We evaluate our AI models by finding meaningful correlations with other customizable features of avatars as well as identity-related demographic information collected from the players.

THEORETICAL FRAMEWORK

Avatars and Identity

Ducheneaut et al. (2009) conducted a qualitative assessment of players' values and avatar customization behaviors for three different videogames. Through self-reported surveys, players described how they customized their avatars, providing insight into preferences for particular visual characteristics. Some players based characters on the notion of "idealized selves" (Lin and Wang 2014) where avatars possessed visual characteristics that players themselves desired in real-life; others created avatars with identities distinctly different from their own in order to conform to societal norms (Dunn and Guadagno 2012). Researchers also discovered that players conformed to role expectations based on their avatars' appearance (e.g., players with taller avatars were more aggressive in negotiations than counterparts with shorter avatars) in a phenomenon known as the Proteus Effect (Yee and Bailenson 2007). Such research shows how avatars can be used to quantitatively study experiences that are often times deemed subjective. This includes real-world social identity constructs such as racial attitudes and stereotypes that manifest in the virtual world (Ash 2015). We previously studied how characters in commercial games and avatars created by players could be computationally analyzed to reveal such social phenomena (Lim and Harrell 2015c).

AI Clustering for Player Modeling

AI clustering is the process of (often automatically) categorizing a large data set of players into a smaller set of groups (clusters). Members within groups generally resemble each other; members between groups are often dissimilar. Resemblance is based on characteristics (e.g., demonstrated proficiency, spending behaviors, etc.) enabling designers to reason more generally about common behavioral patterns. Together, this helps designers to feasibly model diverse groups of players to help improve a game's engagement, user experience, and enjoyment (Yannakakis and Hallam 2009; Yannakakis and Togelius 2011). Many clustering approaches for modeling players exist: Drachen et al. (2014) found that clusters of different algorithms could be distinguished by (1) how interpretable they were, (2) whether they depicted allowed/possible states, (3) how distinct they were from each other, and (4) how well they represented the original data set. In our previous work, we extended analysis beyond numerical data to include images and text (Lim and Harrell 2015a). This paper's chosen clustering techniques (described next) were informed by these previous findings.

Algorithms for Clustering

We used the non-negative matrix factorization (NMF) and archetypal analysis (AA) algorithms for our AI models. Both techniques are algorithmically similar but differ in produced clusters due to different constraints. We start with the following formulation. Assume a player i is represented as a feature vector $x_i \in \mathbb{R}_m$ where m is the number of representation features. Given a data set of n players $V = \{x_1, x_2, ..., x_n\}$, (regular) matrix factorization decomposes the matrix $V \in \mathbb{R}_{n\times m}$ into an approximation $\hat{V} \in \mathbb{R}_{n\times m}$—a product of two smaller matrices $W \in \mathbb{R}_{n\times k}$ and $H \in \mathbb{R}_{k\times m}$. The value k specifies the number of desired clusters. Each (basis) vector $h_j \in H$ defines a cluster and membership requirements; they are often termed *prototypes*. Each weight vector $w_{ij} \in W$ associates each player i with prototype j. Each player is then assigned to the cluster with the highest associated value.

Non-Negative Matrix Factorization

Non-negative matrix factorization is an effective image analysis algorithm, capable of as finding different components of faces (e.g., eyes, mouth, etc.) from facial image data (Lee and Seung 1999) and classifying documents of text (Xu et al. 2003). It imposes non-negativity constraints on both matrices W and H when minimizing $||V - \hat{V}||^2_F$, where

$||X||_F^2 = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} |x_{ij}|^2}$ is the Frobenius norm of the matrix X. The constraints produce basis vectors that differ from other matrix factorization image analysis techniques, which allow negatively-weighted compositions (e.g., eigen-faces) (Sirovich and Kirby 1987).

Archetypal Analysis

Archetypal Analysis (AA) can be viewed as a stricter version of NMF[1]. The prototypes of AA often correspond to actual members in the data set that "stand out" (hence the term: archetypes.) Eugster (2011) demonstrated how AA discovered the four archetypes of "bench-warmer", "rebounder", "three-point shooter", and "offense" in a data set of basketball players defined by their abilities. Each archetype corresponded to top-players for their respective positions. The algorithm works by imposing additional constraints to the matrix factorization. The first set of constraints are that each cluster's prototype $h_j = \sum_{i=1}^{n} \beta_{ij} x_i$ and $\beta_{ij} \geq 0$. The second set of constraints are that weights w_{ij} associating player i with each cluster j all $\sum w_{ij} = 1$. Visually, archetypes lie on the boundary of the data set and form a convex hull. Each player can thus be represented as mixtures of archetypes (Seth and Eugster 2014). AA has been shown to be effective at tasks like image clustering and for recommendation systems (Bauckhage and Thurau 2009; Sifa et al. 2014).

AIRVATAR-RPG: AN AVATAR CUSTOMIZATION SYSTEM

AIRvatar-RPG is a custom avatar creation application[2] set in the context and style of a traditional computer role-playing game (RPG) (Figure 1). The system is integrated with our remote data collection system *AIRvatar* (Lim and Harrell 2015c).

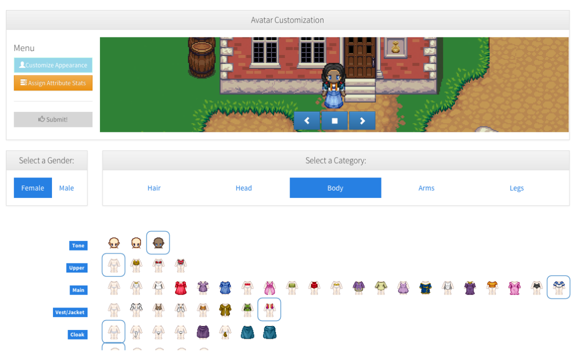


Figure 1: Screenshot image of the user interface of *AIRvatar-RPG*.

Avatar Customization Components

Static Media Assets - Images

Players were free to choose either male or female avatar genders for their character[3], each with a base image and assets for five main categories (hair, head, body, arms, and legs)

1. In fact, we used the convex hull non-negative matrix factorization (CHNMF) algorithm (Thurau et al. 2009) in place of standard AA in this paper. Results are equivalent to AA, but are faster to compute.

2. Art assets from the publicly available *Mack Looseleaf Avatar Creator* (`http://www.geocities.jp/kurororo4/looseleaf/`) and the *Liberated Pixel Cup* (`http://lpc.opengameart.org/`)

3. We follow role-playing game conventions here to study the effects of binary models of gender. However, we recognize distinctions between biological sex and gender, and the other nuanced ways to represent genders of players and avatars. In future work, we seek representations that decouple biological sex and gender.

and sub-categories for more fine-grained options. Each avatar was 32×48 pixels in size, animated in a walkcycle, and viewable in four rotations (front, back, left, and right).

Numerical Attributes -- RPG Stats

Players customized both their character's visual appearance and statistical attributes values of six commonly used videogame attributes on a 7-point scale. Table 1 lists and provides descriptions for each of these attributes. Each attribute was initially defaulted to a value of 4 points, with 3 points unallocated. Players had to allocate all 27 points for their avatar.

Attribute	Description
Strength	Character's ability to deal damage. (i.e., Damage Points)
Endurance	Character's ability to receive damage. (i.e., Health Points)
Dexterity	Character's ability to move quickly and accurately. (i.e., Action Points)
Intelligence	Character's ability to 'level up' quicker. (i.e., Experience Points)
Charisma	Character's ability to charm, convince, or converse well. (i.e., Charm)
Wisdom	Character's ability to cast spells and magic. (i.e., Magic Points)

Table 1: Table detailing the six statistical attributes in *AIRvatar-RPG*.

User Study

We conducted a study with consenting users from the social discussion website *Reddit* (`/r/samplesize`). They were informed that anonymous analytical data would be collected for research purposes. Out of 191 participants, 104 participants (54%) identified as "Male", 81 (43%) identified as "Female", and 6 (3%) listed "Other". 154 participants (80%) were "18-24" years old, 32 (17%) were "25-34" years old, with the rest at $< 1\%$. They also completed a BIG-5 personality test (Gow et al. 2005) and provided demographic information.

METHODS

Image Representations

Each avatar image was cropped from its spritesheet into a 32×48 front-facing image of the avatar (in its idle state). Moving away from our previous efforts in (Lim and Harrell 2015a) that used Red-Green-Blue-Alpha (`RGBA`), we opted to use the Alpha-Hue-Saturation-Brightness (`AHSB`) representation. `HSB` (alternatively termed `HLS`, where `L` is lightness) has historically been considered an improvement over `RGB` (Berk et al. 1982) as it exploits the psychology of color perception. Humans look for variation in hues (color), saturation, and brightness, and can more easily specify a color along those dimensions. In contrast, the `RGB` model is based on physical limitations of the red, green and blue phosphors of color cathode ray tube (CRT) screens. Figure 2 shows an original image in the respective channels.

(a) AHSB (b) A (c) H (d) S (e) B

Figure 2: The images show the different image representations of avatars.

Alpha (`A`) values display transparent areas as black and non-transparent areas as white (Figure 2b); note that avatars do not have semi-transparent parts, and thus the data in the alpha channel is binary. Brightness (`B`) values represent a grayscale version of the base image (Figure 2e). Saturation (`S`) values are normalized to the brightness channel: bright colors like yellow, red, and blue will have higher values (Figure 2d), while white and black areas will have low saturation values (e.g. the white pants in Figure 2d). Hue (`H`) is normalized to these (normalized) `S` values: black (low-value) regions in the `S` image (Figure 2d) correspond to black in the H image (Figure 2c). Blue colors have average hue values and show as gray in Figure 2c (e.g. blue cloak) while red colors have high hue values (e.g. red vest and hair in Figure 2c). The `A`, `H`, `S` and `B` channels can be combined together by creating tuples, each representing a pixel. The `AHSB` version (Figure 2a) is a combination of all channels.

Model Construction

The dataset of $N = 191$ players is represented by the matrix $V = \{v_1, v_2, \ldots, v_N\} \in \mathbb{R}_{N\times M}$. Each $v \in \mathbb{R}_M$ vector represents a player where M is the feature dimensionality—the length of flattened array of a chosen image representation. We constructed models for different combinations of the `A`, `H`, `S`, and `B` channels (e.g., `A`, `AB`, `SB`, `AHB`) in order to investigate the effect of representation on the performance of the clustering algorithms. Hence, the number of features for each model was $M = 32 \times 48 \times C$, where C represents the number of channels required per pixel, which is based on the chosen image representation.

Model Evaluation

We evaluated our models in several ways. The first deals with a qualitative assessment of clusters produced by our models for both clustering algorithms. The second focuses on cluster correlations with the statistical attributes of players' avatars. The third looks at the relationships between the clusters and the genders of both players and their avatars.

Qualitative Analysis of Clusters

Clusters were analyzed qualitatively by visually inspecting the resultant clusters characterized by (1) the base (prototype) image and (2) the members of the dataset that are associated with the clusters. Given the variation of cluster membership sizes, we considered the top-5 weighted individuals during our qualitative evaluation. This revealed how image clusters could be meaningfully interpreted aesthetically and how distinct they were from each other.

Cluster Correlations with Statistical Attributes

We calculated Pearson's correlations between the cluster weights of each player and each of the statistical attributes assigned to their avatar. Doing so enabled us to discover if image clusters—computed without knowledge of alternative avatar representations—could reveal the structures underpinning developers' design of attribute choices and players' values toward attribute distributions. Statistically significant correlations—defined as having $p < .05$ and adjusted for repeated tests—between image clusters and numerical attributes imply that models meaningfully represented both developer and player preferences.

Cluster Associations with Player and Avatar Genders

We investigated if the image clusters obtained were associated with (a) player genders, (b) avatar genders, and (c) player-avatar gender pairings. These associations were computed by constructing cross-tabulations against a player's assigned cluster. Our focus was on whether image clusters could distinguish between either player or avatar genders without prior knowledge. Additionally, we investigated if clusters could identify players who were gender-bending—playing as an avatar with a different gender identity as their own.

RESULTS

Descriptive Statistics of Gender Distributions

Table 2 shows the cross-tabulation of genders of both players and constructed avatars. While we had more male than female players and a relatively small pool of players identifying as "other," we obtained an even number of avatars of both genders. Thus, a larger percentage of male players (18.2%) created avatars of the opposite gender as compared to females (7.4%).

Player Gender	Avatar Gender: Male	Avatar Gender: Female	Total
Female	6	75	81
Male	85	19	104
Other	4	2	6
Total	95	96	

Table 2: Table showing the distribution of both player and avatar genders.

Descriptive Statistics of Attributes

Figure 3 shows the distribution of the statistical attributes of avatars from the data set. Both "strength" and "wisdom" had the lowest means and comparatively larger standard deviations. They featured strongly in our model, as described in the later sections.

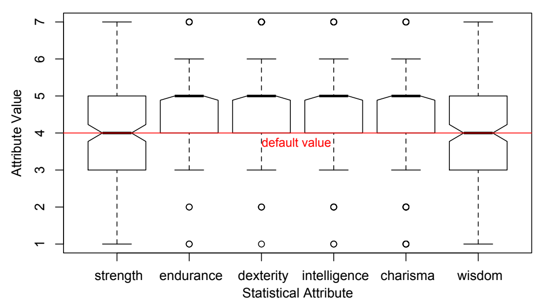


Figure 3: A boxplot showing the distribution of statistical attributes.

Image Clusters: NMF

Figure 4 shows the results of three different NMF models computed based on different image representations. For each image grid, the first column shows the computed base (prototype) image. The next five columns show the top-weighted avatars for each base image. The next five columns show the top-weighted avatars for each base image.

Alpha Channel

The `A` (Figure 4a) NMF model produces clusters that differ from one another based on overall body shape and equipped accessories with high levels of detail. For example, the first cluster is characterized by having a sword equipped—despite the avatars each being visually different in terms of gender, hairstyles, clothings, etc. The other clusters can also be characterized by female hair accessories (row 2), armor-less males (row 3), cloak style (row 4 & row 6), gloves/gauntlets (row 5), shoulder pads (row 7), wings (row 8), under-armored females (row 9) and females with long, straight hair (row 10).

(a) A (b) B (c) AHSB

Figure 4: Images of NMF model results from different image representations. Cluster prototypes isolate specific parts that distinguish between clusters.

Brightness Channel

The `B` (Figure 4b) NMF model clusters differ from one another based on elements within the main body of the avatar. This contrasts with clusters of the `A` model that focus on artifacts mainly outside the body. Again, a high level of detail exists—males with unobstructed faces (row 2), females without shoes (row 6) or different skirt lengths (row 9 and 10).

AHSB Channels

The `AHSB` (Figure 4c) NMF model produces clusters that also account for similarities in color hues for specific portions of the avatar. For example, it discerns the outline of the avatars (based on their alpha) to produce clusters 9 and 10 that show avatars with large armors, males with large armor (row 7) and females without capes/cloaks (row 8). Other clusters include avatars with blue or dark clothes (row 6), avatars with red hair (row 4) or characters with less saturated, earthly tones in their clothes and hair (row 5).

Image Clusters: AA

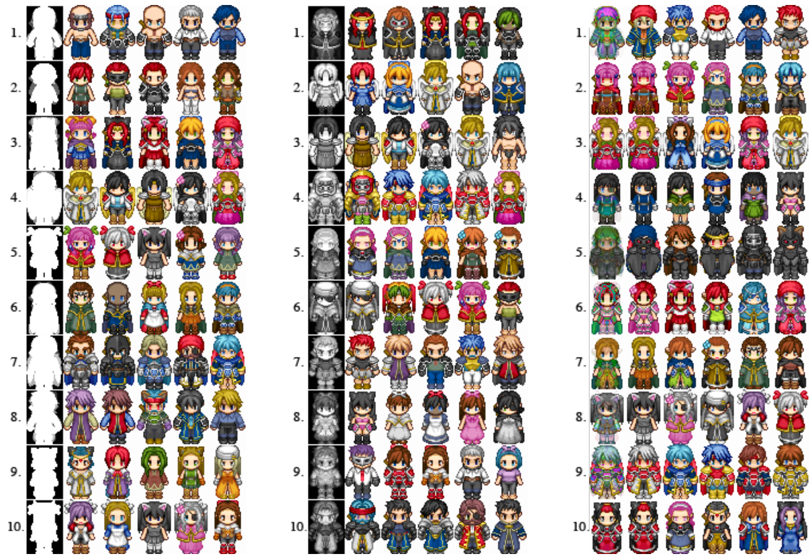
Figure 5 shows the results of AA models computed based on different image representations. The presentation of images is similar to that of the NMF clusters described earlier.

Alpha Channel

The `A` (Figure 5a) AA model constructs clusters that differ from one another based on entire body silhouettes. These silhouettes have distinct identifiable features. E.g., rows 1 and 2 are unarmored males and females respectively; row 3 and 7 are caped females and males respectively; row 4 shows winged avatars. Rows 5, 9, and 10 are females with different hair accessories. Row 6 shows both male and female caped avatars without elaborate headgear.

Brightness Channel

The `B` (Figure 5b) AA model constructs clusters based on more detailed aspects of the avatar images than silhouettes. For example, avatars of row 4 not only have large armors, but they



(a) A (b) B (c) AHSB

Figure 5: Images of AA model results from different image representations. Silhouettes shapes of cluster prototypes distinguish between clusters.

are light-colored, too. Row 2 and 3 both feature avatars with wings but differ based on the hair darkness. The base images all closely resemble the highest-weighted avatar images.

AHSB Channels

The AHSB (Figure 5c) AA model constructs clusters that identify similarities in color hues for different clothing items and body parts of avatars, in addition to overall body shape. For example, this combination of channels discerns between bulky, dark-armored characters in row 5 (from a combination of S, B, and A channels), bright colorful clothing and hair in row 6, and cloaked avatars dressed in earthly tones in row 7, akin to NMF's row 5 of Figure 4c.

Correlations between Clusters and Attributes

Table 3 provides a summary of the best performing NMF and AA models—defined as having **at least** 3 significant correlations at the $p < 0.05$ significance level. Table 4 of the Appendix shows the full cluster-attribute correlations. The convention used for naming models is: <image_channels>_<algorithm>_k<no_of_clusters>. From Table 3, we find that NMF models (`AB-NMF-k6` & `A-NMF-k6`) and AA models (`SB-AA-k3` & `AHSB-AA-k6`) have the highest number of statistically significant correlations with attributes. For each model, we discuss characteristics of the clusters with significant correlations.

NMF Models with High Statistical Attribute Correlations

We found a cluster of `AB-NMF-k6` (Figure 6a) that had significant correlations with statistical attributes; visually, the cluster largely consisted of male avatars with dark costumes and short hair. These avatars were positively correlated with "strength" (.250) and "endurance" (.230) and negatively correlated with "wisdom" (.278) attribute values. The `A-NMF-k6` model also revealed a cluster of significant correlations (Figure 6b); visually, the cluster largely consisted of female avatars with hair accessories. It was positively correlated with "wisdom" (.240) and negatively correlated with "strength" ($-.235$) and "endurance" (.234).

Algorithm	Format	k	STR	END	DEX	CHA	INT	WIS	# Significant Correlations
NMF	`AB`	6	1	1	-	-	-	1	3
NMF	`A`	6	1	1	-	-	-	1	3
AA	`AHSB`	6	2	1	-	-	-	1	4
AA	`HSB`	4	1	1	-	-	-	1	3
AA	`HSB`	8	1	1	-	-	-	1	3
AA	`SB`	3	2	1	-	-	-	2	5
AA	`SB`	7	1	-	-	-	-	2	3

Table 3: Best models ($\geq$ 3 correlations at $p < .05$) between clusters and attributes. Cols 4–9 show the # of sig. correlations associated with that attribute.

(a) Cluster of dark-clothed males with short hair and exposed faces. They had high physical (str/end) & low magical (wis) attributes.

(b) Cluster of bright-haired/clothed females, mostly adorning hair accessories. They had low physical (str/end) but higher magical (wis) attributes.

Figure 6: Separate NMF model clusters with significant attribute correlations.

AA Models with High Statistical Attribute Correlations

We found significant correlations for two clusters in the `SB-AA-k3 model` (Figure 7). Figure 7a shows the cluster which largely consisted of female avatars with saturated, bright clothing; these were negatively correlated with "strength" (.224) and "endurance" (.228) and positively correlated with "wisdom" (.253). Figure 7b shows the other cluster of male avatars with desaturated, dark clothing—mostly dark cloaks—and short hair; these were positively correlated with "strength" (.218) and negatively correlated with "wisdom" (.251).

(a) Cluster of female avatars with saturated, bright clothing. They were associated with low physical (str/end) but higher magical (wis) attributes.

(b) Cluster of male avatars with dark cloaks. They were associated with dealing high damage (str) but having low magical (wis) attributes.

Figure 7: Clusters of model `SB-AA-k3` with significant attribute correlations.

We found significant correlations for two clusters in the `AHSB-AA-k6` model (Figure 8). Figure 8a shows a cluster consisting of female avatars with saturated bright clothing or big, dark hairstyles with bright accessories; these were negatively correlated with "strength" (.278). Figure 7b shows the cluster consisting mostly of male avatars with short, bright hair—mostly with facial hair and dark clothes; these were positively correlated with "strength" (.270) and "endurance" (.251) and negatively correlated with "wisdom" (.341).

(a) Female avatars with big dark hair—mostly with bright hair accessories and two with eye patches. They were associated with low strength attributes.

(b) Cluster of dark-clothed males with shoulder pads, most with face masks/mustaches. They had high physical (str/end) & low magical (wis) attributes.

Figure 8: Clusters of model `AHSB-AA-k6` with significant correlations.

Relationships between Clusters with Player/Avatar Genders

Informed by the four models with high correlations discussed in the previous section, we present results from cross-tabulating the clusters from each model with the gender breakdown of both players and their avatars. For each table listed, the rows correspond to the model's clusters (#rows= k number of clusters) and columns correspond to player–avatar gender pairs (#columns= 3 player –genders $\times 2$ avatar–genders $=$ 6 gender pairs). The tables of cross-tabulations referred to in this section are in the Appendix. Here, we provide descriptive summaries of each of them to highlight characteristics of the resultant clusters.

Player/Avatar Gender Distributions in Clusters of NMF Models

In Table 5 the `A-NMF-k6` model had two clusters that associated with a specific avatar gender—one containing only female avatars and one only male avatars (Figure 9a). Three clusters had diverse associations with both avatar genders. In Table 6, the `AB-NMF-k6` model had four clusters that associated with a specific avatar gender—two clusters contained only male and two other clusters contained only female avatars. Of the two remaining clusters, one cluster had a diverse mix of associated avatar genders (Figure 9b).

(a) Cluster of male-only avatars in **A-NMF-k6**.

(b) Cluster of diverse avatars in **AB-NMF-k6**

Figure 9: NMF clusters that revealed patterns related to avatar gender.

Player/Avatar Gender Distributions in Clusters of AA Models

In Table 7, the `SB-AA-k3` model did not have clusters that associated with a specific avatar gender. However, it can be seen that cluster 1 is often associated with female avatars, cluster 2 with male avatars, and cluster 3 evenly split between male ($n = 36$) and female ($n = 37$) avatars. In Table 8, the `AHSB-AA-k6` model had two clusters which contained only female avatars, including the cluster shown in Figure 8a. It should be noted that the cluster in Figure 8a not only contains only female avatars, but were all created by female players too.

DISCUSSION

The k Effect

The challenge with unsupervised learning approaches lies in determining the optimal number of clusters k. Error measures used during matrix factorization (e.g., Frobenius norm) are by definition optimal when equal to zero, leaving them susceptible to overfitting. While regularization (e.g., cutoff thresholds) may work during matrix factorization, they are not effective for cross-model comparisons. Error measures are always smaller with a higher k; it is thus challenging to find an optimal model with small k without qualitative analysis.

Increasing/Decreasing k

Figure 10 shows the effect of increasing the number of clusters k for both NMF and AA models. Both models' clusters differentiate based on coarser differences—like gender—at low values of k. Differences become more localized and granular as k increases. With brightness, NMF base images isolate specific areas of the image (e.g., faces and armor in Figure 10c); AA base images are avatars with distinct appearance. Similar clusters exist between both algorithms, both general (e.g., avatars with wings) and specific (e.g., smaller-sized female avatars in the red dress and purple inner-wear). A high k helps with identifying differences and NMF models are easier to visually interpret and differentiate qualitatively.

(a) B-NMF-k2 (b) B-NMF-k4 (c) B-AA-k2 (d) B-AA-k4

Figure 10: Images of clusters with varying k. At lower k, differences are more "global" in scope; At higher k, differences are more localized and granular.

Determining Model Optimality with k Clusters

The notion of an optimal k depends on the goals of AI clustering: our goal was to produce base (prototype) images that can be (a) easily interpreted, (b) result in clusters that were distinguishable, (c) showed correlations with attribute stats, and (d) showed relationships with player/avatar gender mappings. For these goals, an NMF model with k=6 clusters worked well. While the AA model with k=3 clusters had the highest number of correlations, we decided that k=6 provided a slight trade-off in that regard while better describing the relationships between clusters and player/avatar gender. Hence, the optimal number of clusters was k=6 for both NMF and AA and is dependent on the metrics for defining optimality.

NMF vs. AA

Effectiveness of Different Image Representation Formats

The NMF models with the highest number of attribute correlations used `A` and `AB` image representations, while AA models with the highest number of attribute correlations used `SB` and `AHSB` image representations. This implies that NMF works better without hue and saturation, while conversely, AA works better *with* hue and saturation. This highlights an important difference between both algorithms and their effectiveness in different situations.

Human vs. Machine Interpretation

AA produced prototype images that were valid (i.e. actual individuals in the data set) while NMF produced "parts-based" prototype images (i.e. not valid). The NMF base images however are easier for humans to interpret features that distinguish clusters from each other, particularly at high k values. AA base images simply show us the "archetype" of clusters that possesses the distinguishing feature but requires looking at other members to isolate the distinguishing features. Despite being less human interpretable, the AA models appear to produce clusters that have more significant correlations, indicating that the computer (machine) interprets images different from visual (human) perception.

Implicit Categorization of Statistical Attributes

Clusters and Gender/Role Stereotyping

In our models, the three common attributes were "strength", "endurance", and "wisdom". It highlights how players are inclined toward well-known "physical-fighter" versus "wise-mage" roles and associated stereotypes. For (1) role stereotypes: "strength" and "endurance" were always positively correlated with each other and negatively correlated with “wisdom”. For (2) gender-stereotypes: male avatar-associated clusters were mostly positively correlated with being physically-strong (but unwise); female avatar-associated clusters were mostly positively correlated with being physically-weak (but wise). While we did not separate clusters by player genders here, such stereotyping has been previously observed to be assymetric (i.e. exhibited by players of a certain gender) due to the context, genre, and style of the game (Huh and Williams 2010). We later discuss plans to study this for future work.

Other Attribute Correlations

We found a single occurrence of a model with statistically significant correlations besides the three common ones (strength, endurance, wisdom). In row 15 of Table 4 of the Appendix, a cluster showed negative correlation (.243) with the “charisma” attribute. Avatars in this cluster adorned bright clothing or hair colors with higher-weighted individuals adorning large, bright armors (Figure 11). We reason that the large armor (more intimidating) outweighed colorful appearances (more inviting) to result in lower associations with “charisma”.

Figure 11: Cluster images of avatars negatively correlated with “charisma”

LIMITATIONS & FUTURE WORK

Beyond Visuals and Numbers: Relationships with Text

In *AIRvatar-RPG*, players also input a list of tags and freeform text to describe their avatars. This text-based input can be converted into a bag-of-words representation and clustered in a process called topic modeling. We aim to explore relationships between image and text clusters—revealing how visual clusters may be associated with themes that players use to describe their avatars. This would provide an alternative lens into player values.

Filtering Data by Different Player Demographics and Profiles

We aim to investigate the effects that players’ demographic profiles have on our approaches, achieved by filtering our data set separately based on player gender. More players would be needed to balance the gender ratios. This would enable us to discover if the clusters can identify or reveal effects of gender-bending or other identity-related phenomena like the Proteus effect (Yee and Bailenson 2007), where players conform to role expectations of their avatar regardless of their own identity. From players’ BIG-5 personality data, preliminary results of our approaches showed some cluster correlations, but requires further study. We could also look into customizable avatar personalities like those used in *The Sims*[4].

Clusters for Design and Generation

We have shown effective image analysis over our previous endeavors (Lim and Harrell 2015a). We hope to evaluate it on avatars with different graphical styles including the analysis of 3-dimensional (3D) characters. Being able to robustly produce clusters with more clearly defined membership requirements would enable procedural assessment of game characters that do not rely on—and not constrained by—the underlying representations of characters. This could improve gameplay elements that dynamically adapt to players based on more human perception of visual similarities as in (Lim and Harrell 2015b).

4. http://www.thesims.com

CONCLUSIONS

Players spend large amounts of time customizing their avatars in videogames. These avatars are more than visual artifacts on a computer screen. They reflect the values and preferences of both the developer (through the system design and customization options made available) and the player (through the customization choices made). We showed how AI clustering algorithms, namely NMF and AA, could cluster avatar images automatically based on concretely defined (e.g., avatar gender) as well as more subjective (e.g., aesthetic preferences for particular colors, items, or accessories) features. Clusters showed statistically significant correlations with players' preferences for different attributes, e.g., male avatars in dark clothes having high physical attributes (e.g., strength, endurance) but low magical abilities, while female avatars were commonly associate with less physical attributes like magical abilities (e.g., wisdom). We showcased AI as an effective tool for analyzing videogame avatars for developing models of social identity and aesthetic preferences. Without needing knowledge of the underlying implementation of systems, such emergent patterns provide better perspectives of players that are less constrained by predefined design assumptions.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their valuable feedback. This material is based upon work supported by the National Science Foundation under Grant No. 1064495. We also acknowledge the support of a QCRI-CSAIL Collaboration Grant.

APPENDIX

	model	pair	c
1	`AB-NMF-k6`	cluster5-str	0.250
2	`AB-nmf-k6`	cluster5-end	0.230
3	`AB-NMF-k6`	cluster5-wis	-0.278
4	`AHSB-AA-k6`	cluster2-str	-0.278
5	`AHSB-AA-k6`	cluster4-str	0.270
6	`AHSB-AA-k6`	cluster4-end	0.251
7	`AHSB-AA-k6`	cluster4-wis	-0.341
8	`A-NMF-k6`	cluster3-str	-0.235
9	`A-NMF-k6`	cluster3-end	-0.234
10	`A-NMF-k6`	cluster3-wis	0.240
11	`HSB-AA-k4`	cluster1-str	0.221
12	`HSB-AA-k4`	cluster1-end	0.221
13	`HSB-AA-k4`	cluster1-wis	-0.249
14	`HSB-AA-k8`	cluster8-end	-0.254
15	`HSB-AA-k8`	cluster5-cha	-0.243
16	`HSB-AA-k8`	cluster8-wis	0.236
17	`SB-AA-k3`	cluster1-str	-0.224
18	`SB-AA-k3`	cluster2-str	0.218
19	`SB-AA-k3`	cluster1-end	-0.228
20	`SB-AA-k3`	cluster1-wis	0.253
21	`SB-AA-k3`	cluster2-wis	-0.251
22	`SB-AA-k7`	cluster4-str	0.242
23	`SB-AA-k7`	cluster4-wis	-0.258
24	`SB-AA-k7`	cluster5-wis	0.236

Table 4: Table of clusters and attributes with ≥ 3 sig. correlations ($p < .05$)

Cross-Tabulations of Gender Pairs and Clusters in both NMF & AA Models

		Male Avatar			Female Avatar		
		F	M	O	F	M	O
Cluster	1	0	16	0	6	1	0
	2	2	13	2	3	1	0
	3	0	0	0	13	3	0
	4	0	1	0	17	8	2
	5	0	2	1	36	6	0
	6	4	53	1	0	0	0

Table 5: `A-NMF-k6`

		Male Avatar			Female Avatar		
		F	M	O	F	M	O
Cluster	1	0	1	0	2	1	0
	2	2	37	2	0	0	0
	3	1	19	0	8	3	0
	4	0	0	0	63	15	2
	5	3	28	2	0	0	0
	6	0	0	0	2	0	0

Table 6: `AB-NMF-K6`

		Male Avatar			Female Avatar		
		F	M	O	F	M	O
Cluster	1	0	8	0	37	3	1
	2	4	45	2	9	1	0
	3	2	32	2	29	15	1

Table 7: `SB-AA-k3`

		Male Avatar			Female Avatar		
		F	M	O	F	M	O
Clusters	1	0	1	0	27	11	2
	2	0	0	0	5	0	0
	3	1	2	0	38	6	0
	4	3	44	1	3	1	0
	5	0	0	0	1	1	0
	6	2	38	3	1	0	0

Table 8: `AHSB-AA-k6`

BIBLIOGRAPHY

Ash, Erin. 2015. "Priming or Proteus Effect? Examining the Effects of Avatar Race on In-Game Behavior and Post-Play Aggressive Cognition and Affect in Video Games." *Games and Culture.*

Bauckhage, C., A. Drachen, and R. Sifa. 2014. "Clustering Game Behavior Data." *Computational Intelligence and AI in Games, IEEE Transactions on* PP (99): 1–1. ISSN: 1943-068X.

Bauckhage, Christian, and Christian Thurau. 2009. "Making archetypal analysis practical." In *Pattern Recognition,* 272–281. Springer.

Berk, Toby, Arie Kaufman, and Lee Brownston. 1982. "A Human Factors Study of Color Notation Systems for Computer Graphics." *Comm. of the ACM* 25 (8): 547–550.

Boellstorff, Tom. 2008. *Coming of age in Second Life: An anthropologist explores the virtually human.* Princeton University Press.

Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things Out: Classification and its Consequences.* MIT Press.

Drachen, Anders, Christian Thurau, Rafet Sifa, and Christian Bauckhage. 2014. "A comparison of methods for player clustering via behavioral telemetry." *arXiv preprint arXiv:1407.3950.*

Ducheneaut, Nicolas, Ming-Hui Wen, Nicholas Yee, and Greg Wadley. 2009. "Body and Mind: A Study of Avatar Personalization in Three Virtual Worlds." In *Proc. of the 27th SIGCHI Conference on Human Factors in Computing Systems,* 1151–1160. USA.

Dunn, Robert Andrew, and Rosanna E. Guadagno. 2012. "My avatar and me: Gender and personality predictors of avatar-self discrepancy." *Computers in Human Behavior* 28 (1): 97–106.

Eugster, Manuel JA. 2011. "Archetypal athletes." *arXiv preprint arXiv:1110.1972.*

Gee, James Paul. 2003. "What video games have to teach us about learning and literacy." *Computers in Entertainment (CIE).*

Gow, Alan J, Martha C Whiteman, Alison Pattie, and Ian J Deary. 2005. "Goldberg's 'IPIP' Big-Five factor markers: Internal consistency and concurrent validation in Scotland." *Personality and Individual Differences* 39 (2): 317–329.

Harrell, D. Fox. 2009. “Computational and Cognitive Infrastructures of Stigma: Empowering Identity in Social Computing and Gaming.” *Proceedings of the 7th ACM Conference on Cognition and Creativity:*49–58.

Huh, Searle, and Dmitri Williams. 2010. “Dude looks like a lady: Gender swapping in an online game.” In *Online worlds: Convergence of the real and the virtual,* 161–174.

Lakoff, George. 1990. *Women, Fire, and Dangerous Things.* Cambridge University Press.

Lee, Daniel D, and H Sebastian Seung. 1999. “Learning the parts of objects by non-negative matrix factorization.” *Nature* 401 (6755): 788–791.

Lim, Chong-U, and D. Fox Harrell. 2014. “Developing Social Identity Models of Players from Game Telemetry Data.” In *Proc. of the AAAI AIIDE’14 Conference,* 125–131.

———. 2015a. “Developing Computational Models of Players’ Identities and Values from Videogame Avatars.” In *Proc. of 10th International Conference on FDG’15.*

———. 2015b. “The Marginal: A Game for Modeling Players’ Perceptions of Gradient Membership in Avatar Categories.” In *Proc. of the 2nd AIIDE Workshop on EXAG.*

———. 2015c. “Toward Telemetry-driven Analytics for Understanding Players and Their Avatars in Videogames.” In *CHI’15 Extended Abstracts on Human Factors in Computing Systems,* 1175–1180. S. Korea.

Lin, Hsin, and Hua Wang. 2014. “Avatar creation in virtual worlds: Behaviors and motivations.” *Computers in Human Behavior* 34:213–218.

Rosch, Eleanor. 1999. “Principles of categorization.” *Concepts: Core readings:*189–206.

Seth, Sohan, and Manuel J. A. Eugster. 2014. *Probabilistic Archetypal Analysis.* Technical report. arXiv.org. `http://arxiv.org/abs/1312.7604`.

Sifa, Rafet, Christian Bauckhage, and Anders Drachen. 2014. “Archetypal Game Recommender Systems.” *Proc. KDML-LWA.*

Sirovich, Lawrence, and Michael Kirby. 1987. “Low-dimensional procedure for the characterization of human faces.” *Josa a* 4 (3): 519–524.

Thurau, Christian, Kristian Kersting, and Christian Bauckhage. 2009. “Convex non-negative matrix factorization in the wild.” In *Proc. of the IEEE ICDM’09.* 523–532. IEEE.

Xu, Wei, Xin Liu, and Yihong Gong. 2003. “Document clustering based on non-negative matrix factorization.” In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval,* 267–273. ACM.

Yannakakis, Georgios N, and John Hallam. 2009. “Real-time game adaptation for optimizing player satisfaction.” *Computational Intelligence and AI in Games, IEEE Transactions on* 1 (2): 121–133.

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

Yee, Nick. 2006. “The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments.” *Presence* 15 (3): 309–329.

Yee, Nick, and Jeremy Bailenson. 2007. “The Proteus effect: The effect of transformed self-representation on behavior.” *Human communication research* 33 (3): 271–290.