Comparing Clustering Approaches for Modeling Players' Values through Avatar Construction

Chong-U Lim and **D. Fox Harrell**
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA
{culim, fox.harrell}@mit.edu

Abstract

Videogame avatars provide an expressive avenue for players to represent themselves virtually. Research has shown that these avatars, while virtual, can reveal aspects of players' identities, along with physical, social, and cultural values of the real-world. In this paper, we present an approach for modeling player values through their avatars using artificial intelligence (AI) clustering techniques. In a study with 191 participants who created avatars using our system, we provide a thorough comparison of the techniques across numerical, textual, and visual data. Our findings showed that these data structures can effectively reveal players' values and preferences, such as conforming to stereotypes of character roles using statistical attributes, modeling nuances in text descriptions of avatars, and identifying "best-example" (prototypical) avatar appearances that players can be quantitatively shown to conform to. Our findings suggest that AI clustering approaches can be used to model players to yield insight into implicitly held values in a data-driven manner through virtual avatars.

Introduction

Many videogames provide avatar constructors for players to customize their virtual identity representations. The common computational data structures used for such representations include images, text, numerical data, or procedural behavioral rules (Harrell 2009). While technically implemented as virtual identities, they should not be viewed simply as results of a user-directed creation process (Yee et al. 2011). In research highlighting the relationship between real and virtual identities, virtual identities can: (1) reveal aspects of a player's real-world identity like demographic information, personalities, and motivations (Tekofsky et al. 2013; Canossa, Martinez, and Togelius 2013), (2) reveal phenomena reflecting real-world social constructs (e.g., notions of ideal body types or stereotypes) (Dunn and Guadagno 2012; Harrell and Veeragoudar-Harrell 2012), and (3) affect players' real-world behaviors and performance (Yee and Bailenson 2007; Steele and Aronson 1995).

In this paper, we present an approach to modeling player values through their avatars using artificial intelligence (AI) clustering techniques. We compare four different AI clustering techniques across four types of data structures commonly used to implement virtual avatars. Using such data-driven AI approaches enables us to gain understanding of the values and preferences held by players in a "bottom-up," emergent manner. We also seek for our findings to provide a comparative summary of the performance of the different techniques across different types of data structures.

Background

Here we cover related research on player modeling, avatars and identity, cognitive categorization, and AI clustering.

Player Values and Blended Identities

To study these values and avatars computationally, we use Harrell's concept of a "blended identity" (Harrell 2013), where some aspects of a player's real-world identity (e.g., values, preferences, appearance, understanding of social categories, etc.) are projected onto the actual implemented avatars. To formally describe the data structures used to construct these virtual identities (avatars), which are highly expressive in visual appearance and behaviors, we use Harrell's taxonomy of technical components identified for computational identity technologies (Harrell 2009). They are 1) static media assets, 2) flat text profiles, 3) modular graphical models, 4) statistical/numerical representation, 5) formal annotation, and 6) procedural/behavioral rules. Combining this taxonomy with cognitive science theories defining social relationships, preferences, metaphors, and values, we can begin to computationally model aspects of players' values and preferences categorization phenomena (Harrell 2009).

Cognitive Categorization and Prototypes

A thorough discussion of the literature from cognitive science and the sociology of classification is beyond the scope of this paper. However, to contextualize our motivations of modeling players' avatars with clustering, we use cognitive categorization theories fron (Rosch 1999) that identities **prototypes** as members that are "better examples" of a category than others. Categorization is based on the distances relative to these prototypes, termed as "centrality gradience" in (Lakoff 1990). This notion of centrality to prototypes is what motivates the use of AI clustering in this paper.

AI Clustering Approaches

While existing research based on cluster-based or emergent player models exist (Drachen et al. 2012; Drachen, Canossa, and Yannakakis 2009), we focus on several prototype-based categorization AI clustering techniques. We omit a details of both principal component analysis (Jolliffe 2005) and K-means clustering (MacKay 2003) given their general familiarity. While PCA, NMF, and AA are applicable beyond clustering, we adopt the view of (Drachen et al. 2014) who applied them for behavioral clustering to identify different *World of Warcraft* types based on level progression. They found that clusters differed based on interpretability, distinction from one another, the depiction of legal/possible representations, and how representative of the data set they were. Our approaches differ as (1) we use not just numerical data, but also textual (free-text descriptions) and visual (images) data, and (2) we focus on modeling aspects of users' real-world values (e.g., implicit categorization of RPG classes, preference for different types of stories in text descriptions), which provided insight beyond in-game performance.

Non-negative Matrix Factorization Non-negative matrix factorization (NMF) is an algorithmic process for representing data as a combination of derived factors, each representing distinct "parts"-based representations. Formally, given a data set of points $V = \{x_1, x_2, ..., x_n\}$, NMF decomposes the matrix $V \in \mathbb{R}_{n \times m}$ into an approximation $\hat{V}$ – the product of two matrices $W \in \mathbb{R}_{n \times k}$ and $H \in \mathbb{R}_{k \times m}$, with $v_{ij} \in V$, $w_{ij} \in W$, and $h_{ij} \in H$ all ≥ 0. The value k specifies the number of parts desired $k \ll n, m$. NMF minimizes $||V - \hat{V}||_F^2$, where $||X||_F^2 = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{m} |x_{ij}|^2}$ is the Frobenius norm of the matrix X. Each row in matrix H is an m-dimension *basis vector* and each column in matrix W relates each sample in v_i to each basis vector via *coefficients* $w_{ij} \in W$, describing contribution of a basis vector j in sample x_i. Besides image analysis, like identifying parts of faces (e.g., eyes, mouth, etc.) from facial images (Lee and Seung 1999), NMF has been effective in other areas like procedural content generation (Shaker and Abou-Zleikha 2014).

Archetypal Analysis Archetypal Analysis (AA) (Cutler and Breiman 1994), is a method for reducing the dimensionality of multivariate data. Given a set of multivariate data points, the aim of AA is to be able to represent each data point as a *convex combination* of a set of key data points called **archetypes**. For example, applying AA on a dataset of basketball players statistics (Eugster 2011) revealed the four archetypes of"benchwarmer," "rebounder," "three-point shooter," and "offensive." Individual players in the data set was then represented as a hybrid mixture of the archetypes (Seth and Eugster 2014). Formally, given a data set of points $\{x_1, x_2, ..., x_n\}$, AA seeks to find a set of archetypes $\{z_1, z_2, ..., z_k\}$ where $z_j = \sum_{i=1}^{n} \beta_{ij} x_i$. Each data point x_i is represented in terms of the k archetypes as $\hat{x}_i = \sum_{j=1}^{k} \alpha_{ji} z_j$. The objective function minimizes the residual sum of squares $RSS = ||x_i - \sum_{j=1}^{k} \alpha_{ij} z_j||^2$ under constraints that weights $\sum \beta_{ij} = 1$ $\beta_{ij} \geq 0$ and coefficients $\sum \alpha ji = 1$ $\alpha_{ji} \geq 0$. Archetypes are located on the convex hull (Cutler and Breiman 1994) and convex mixtures of the data for easier interpretation (Bauckhage and Thurau 2009).

Systems & Applications

We provide an overview of the systems we developed for players to construct avatars and for data to be collected.

AIRvatar and AIRlib

AIRvatar is our analytics system, which collects avatar customization and telemetry data. Underlying *AIRvatar* is the AI toolkit *AIRlib* responsible for transforming the aggregated data, analyzing the data, and visualizing the results. The deployed version of *AIRvatar* uses Javascript and interfaces with the current version of *AIRlib*, which is implemented in Python using the `scikit-learn` (Pedregosa et al. 2011) and Python Matrix Factorization `PyMF`[1] libraries. *AIRlib* uses the Convex Hull Non-negative Matrix Factorization (CHNMF) implementation (Thurau, Kersting, and Bauckhage 2009) in place of AA, which gives equivalent results to AA, but is computationally faster in performance.

Case-study: Heroes of Elibca

We developed an avatar customization system set in the context and style of a traditional computer role-playing game (RPG) called *Heroes of Elibca*. Resources and assets were from publicly available sources (Liberated Pixel Cup 2015; Mack Looseleaf Creator 2015) and players were introduced to the fantasy setting their avatars at the beginning.

Static Media Assets - Images Players chose either male or female avatar genders[2], each with a base image and assets for five main categories (hair, head, body, arms, and legs) and sub-categories for more fine-grained options. Each avatar was $32x48$ pixels in size and could be animated and seen in four rotational views. Here, we only analyzed the $32x48$ front-facing image for each avatar.

Text Profiles – Tags and Descriptions Players were provide two type of text-based representations for their created avatars: (1) a list of word *tags* (e.g., "strong, clever, brooding",) and (2) free-prose text for more verbose descriptions. Both were optional and had examples to guide players.

Numerical Attributes – RPG Stats Players customized both their character's visual appearance and statistical attributes values of six commonly used videogame attributes (strength, endurance, dexterity, intelligence, charisma, and wisdom) on a 7-point scale. Each was defaulted to 4 points with 3 remaining points unallocated for a total of 27 points.

[1]http://code.google.com/p/pymf

[2]We follow role-playing conventions here, but recognize the distinction between gender and sex. In future work, we seek representations that decouple biological sex and gender.

Methods

Model Construction

We represent the avatar of a player i and component j as vector v_{ij}, modeling each dataset of $N = 191$ players with an $N \times M$ matrix V_j. The feature vector v_{image} was created by flattening the 32×48 avatar image. With each pixel represented by RGBA values, we had $M = 32 \times 48 \times 4 = 6144$. For, $v_{attributes}$, we had $M = 6$ corresponding to the six types of attributes. Both v_{tags} and $v_{description}$ were constructed using bag-of-words (BOW) representations. For tags, M=397 was the number of unique tags observed across the data set. For descriptions, M corresponded to the number of unique English word terms observed after tokenizing each of the text descriptions in the data set. It was capped at a default value of M=500. In our experiments, we used k=4 prototypes for each of the technical components except for the numerical attributes, for which we used k=3. These numbers were based off previous results in (Lim and Harrell 2015a; 2015c) that showed that resultant NMF and AA models had sufficiently distinct prototypes for interpretation.

Data Collection

We conducted a user study, approved our institution's human subjects research committee, with consenting participants from the social news and discussion site *Reddit* (`/r/samplesize`). Participants were informed that anonymous analytical data would be collected during customization for research. Out of 191 participants – 104 participants (54%) identified as "Male", 81 (43%) identified as "Female", and 6 (3%) listed "Other." 154 participants (80%) were between "18-24" years old, 32 (17%) were between "25-34" years old, and the other age groups were $< 1\%$.

Results & Analysis

We here present the results from the clustering algorithms on the various technical components of players' avatars.

Statistical Attributes

Table 1 below shows the results of the four clustering algorithms on the numerical statistical attributes. We describe several notable characteristics of the resultant prototypes.

Validity/interpretability of prototypes Results from PCA were distinctly different from the rest, with negative values, as they are eigenvectors (i.e., directions) and not points in the same feature space. Some degree of intepretation can be performed by indirectly using the relative signs of the attribute values. Identifying prototypes is more difficult, requiring the consideration of both highest and lowest-scoring individuals, as signs can be flipped. For example, P1 of NMF could be viewed as a high-strength, low-wisdom prototype, or a high-wisdom, low-strength prototype. In contrast, prototypes from NMF, AA, and K-means were easily interpretable, as they exist in the same feature space. Only the prototypes of AA and K-means were all valid ($1 \leq x \leq 7$), since $P1$ of NMF had an value of 7.5 for "Wisdom". Since the rest of the NMF prototypes are valid, a solution might be to clamp the values.

	C	D	E	I	S	W
P1	1.2	2.0	1.7	3.2	0.2	**7.5**
P2	0.1	4.2	4.4	3.0	**5.3**	0.4
P3	**6.7**	2.2	1.8	2.6	1.8	0.2

(NMF)

	C	D	E	I	S	W
P1	1.2	**6.7**	**6.7**	4.9	6.6	1.0
P2	5.4	4.3	3.3	5.7	1.2	**7.0**
P3	**7.0**	4.0	**7.0**	1.0	**7.0**	1.0

(AA)

	C	D	E	I	S	W
P1	0.2	0.2	0.3	0.1	**0.5**	0.7
P2	**0.9**	0.1	0.2	0.0	0.1	0.4
P3	0.1	0.5	**0.4**	0.6	**0.4**	0.3

(PCA)

	C	D	E	I	S	W
P1	4.3	**5.6**	5.1	4.8	**5.5**	1.8
P2	4.3	4.3	**4.6**	**4.7**	4.5	4.6
P3	4.8	4.7	3.8	5.2	2.6	**5.8**

(K-means)

Key:
***C**harisma **D**exterity **E**ndurance **I**ntelligence **S**trength **W**isdom*

Table 1: The results from each of the clustering algorithms on statistical attributes. Bold values indicate the highest attribute value of a prototype. Values in red are negative.

Number of maximized attributes per prototype Some of the prototypes from AA have more than one attribute that was maximized. Each prototype for the other algorithms only one attribute being maximized. Prototypes in K-means, however, did have prototypes that possessed multiple attributes with values close to the maximum. This shows that NMF and PCA produce prototypes that differ in a one key attribute, while AA and K-means produces prototypes that takes multiple attributes into account.

Degree of similarity (or difference) between prototypes Unlike the others, prototypes from K-means do not vary drastically from one another and appear well-balanced across attributes. This reflects the nature of the K-means algorithm, which seeks prototypes that generalize well to individuals located close to them. Contrastingly, NMF seeks to discover parts-based representations, PCA seeks orthogonal prototypes, and AA are based on extremal individuals, likely resulting in prototypes that differ greatly from one another.

Implicit RPG character roles and categories We categorized each prototype according to familiar RPG roles. We used the highest and lowest-valued attributes of each prototype to make an informed interpretation of each prototype. Results are shown in Table 2. It again highlights the difficulty of identifying and interpreting PCA prototypes, as P1 of PCA appears to be the inverse of P1 of NMF and P3 of K-means, while P2 of of PCA appears similar to P3 of NMF. While the prototypes reflect well-known roles and associated stereotypes, such as weak mages and uncharismatic fighters, we note unusual roles from prototypes in AA, such as a Charismatic Tank. This is unsurprising as these prototypes (archetypes) correspond to actual player creations and

alg.	p.	High	Low	Interpreted Role/Category
NMF	1.	W	S	Weak Mage
	2.	S	C	Uncharismatic Fighter
	3.	C	W	Charismatic Thief
AA	1.	DE	W	Agile Thief
	2.	W	S	Weak Mage
	3.	CES	WI	Charismatic Tank
PCA	1.	S	W	Non-magic Fighter
	2.	C	W	Charismatic Thief
	3.	ES	I	Unintelligent Tank
KM	1.	D	W	Agile Thief
	2.	I	CD	Balanced
	3.	W	S	Weak Mage

Table 2: Table showing the roles of each prototype, categorized by interpreting them using the highest and lowest-valued attributes. Traditional RPG classes are used here.

might reflect a players' intention to create one to "stand out".

Tags

Table 3 shows the resultant prototypes from the four clustering algorithms. For each, the five highest occurring words aggregated from its top five weighted individuals are shown.

alg.	p.	tags *(Sorted by decreasing frequency)*
NMF	1.	intelligent agile quiet adventurous slim
	2.	strong powerful tough leader serious
	3.	smart kind cunning charming charismatic
	4.	clever independent wise quick agile
AA	1.	quick strong shy swordsman accurate
	2.	smart serious charming unglamorous gruff
	3.	clever independent strong wise prepared
	4.	clever charismatic powerful confident brave
PCA	1.	strong powerful leader serious proud
	2.	clever agile intelligent strong empathetic
	3.	intelligent quiet healer long-range ambitious
	4.	quick smart strong cunning self-motivated
KM	1.	strong intelligent fast <blank> narcissistic
	2.	strong blunt confident honest power
	3.	quiet smart short-tempered gruff forceful
	4.	quiet clever intelligent detached deliberate

Table 3: Table showing the prototypes obtained from clustering on tags. The top five-weighted tags are shown.

NMF prototypes are distinct from each other and are each thematically consistent As previously reported in (Lim and Harrell 2015a), the NMF tags are disjoint *between* prototypes, but consistent *within* prototypes. For example, P1 describes an intelligent, agile, and quiet character (e.g., stealthy thief) and P2 describes a strong, powerful, and tough leader (e.g., tank or fighter.) While no tags overlap, some synonymic similarities appear between P1 and P4. It is worth noting that definitions of attributes like "intelligence," may be conflated between common in-game meanings our system used (e.g., for leveling up) and non game-related everyday meanings (e.g., possessing high intellect.)

AA prototypes are fairly distinct from each other, with overlapping tags, but may possess unusual themes The tags from the prototypes of AA appear, like NMF, to be disjoint from each other. P3 and P4 appear relatively particularly similar, both with "clever" as the highest occurring tag. However, an interesting observation is how tags within each prototype are not necessarily thematically consistent. For example, P1 describes a quick and strong, but shy swordsman, P2 describes a charming but gruff character. The reason for this may lie in the fact that (1) tags of each prototype correspond exactly to the list of tags made by players and (2) these prototypes are extremal individuals (archetypes) that represent extremely unique player constructed avatars.

PCA prototypes are fairly distinct from each other, with overlapping tags, and are thematically consistent The prototypes of PCA appear very disjoint, P1 focuses on strength, P2 on intelligence and agility, P3 on intelligent and long-range capabilities, and P4 on speed and cunning. However, there are overlaps between tags used between prototypes. This is somewhat surprising, given the intuition that PCA would maximize orthogonality between prototypes by having distinct tags. But it is worth noting that these overlapping tags do not share the same importance (unlike the "clever" tag of P3 and P4 of the AA prototypes.) The tags within each prototype appear to be thematically consistent.

K-means prototypes are not distinct from each other, with overlapping tags, and possess both consistent and unusually represented prototypes The prototypes of K-means can be separated into two broad categories – strong (P1,P2) vs. (quiet) characters, but there is significant overlap between prototypes. Additionally, some of the prototypes appear to be thematically inconsistent, such as P3 being quiet, but short-tempered and gruff and P4 being both detached and deliberate. Our guess is that since K-means attempts to find centroids within the data set in forming clusters, each prototype appears as an average of individuals around them. Thus, prototypes end up being aggregated mixtures of individuals of the data set, reducing distinctions.

Text Descriptions

Table 3 below shows the results of analyzing the free-text descriptions of avatars using the four clustering algorithms.

NMF description prototypes are distinct from one another in terms of words used, but bear some synonymic similarities regarding story location and setting We observe that sets of words between prototypes appear disjoint, but have synonymic similarities such as "king" (P1) and "chief" (P4), and "town" and "village". Within each prototype, there is a degree of thematic consistency such as P1 about a kingdom setting, P2 about life in the family and home, P3 about mages and magic, and P4 on a farm/town setting (mentioned in the opening of *Heroes of Elibca*.) NMF performs fairly well with fine-grain differences in text.

AA description prototypes are distinct from one another in both words and themes, but top-weighted individuals share high similarities with prototypes. We note firstly

alg.	p.	words *(Sorted in decreasing weights)*
NMF	1.	village family kingdom young king
	2.	town life home help family
	3.	magic light mage parents years
	4.	farm good town land chief
AA	1.	family town help earth like
	2.	food knew village avatar life
	3.	saw months acts days soldier
	4.	light magic force years parents
PCA	1.	town farm family village good
	2.	village good farm chief army
	3.	village family kingdom king food
	4.	close people life like combat
KM	1.	girl home world adventuring gain
	2.	<*blank*>
	3.	goes thief work non career
	4.	like best seek hot soldier

Table 4: Table showing factors obtained from non-negative matrix factorization on the bag-of-words representation of tokenized textual descriptions with varying k. The five highest weighted attributes are shown for each factor.

that each prototype from AA consists of words, which all appear in the description of its top-weighted individual. Thus, each prototype contains words that are highly thematically consistent. We note that there is no overlap in description words used between prototypes and they each focus on fairly disjoint themes. This is a result of finding archetypal player descriptions that are deemed on the extremities of the data set. However, we note that despite this, the words of the prototype do occur frequently in the rest of the top-weighted individuals. It suggests that AA is effective in capturing mixtures of individuals even with high dimensionality data such as tokenized free-text descriptions as demonstrated here.

PCA description prototypes are similar in both actual words used and themes We observe that "village" appears across three different prototypes, with it having the highest weight in both P2 and P3. The themes share similarities with one another, and some of the prototypes bear resemblance to those seen in NMF (e.g., NMF P1 and PCA P3.) Within each description prototype, the words are thematically consistent. The high number of overlapping words between prototypes suggest that a larger number of specified clusters k is required develop more distinct prototypes.

K-means description prototypes are highly disjointed from one another and produced only single top-weighted individuals Prototypes from K-means appear highly disjointed, even turning up a prototype that had no description. While also having distinct prototypes, AA's prototypes had multiple top-weighted individuals, while each K-means prototype had words that appeared only in the single, top-weighted individual. Thus, the K-means prototypes do not effectively model mixtures of text descriptions. We suspect that K-means requires a larger number of iterations for an optimal convergence compared to the other algorithms.

Images

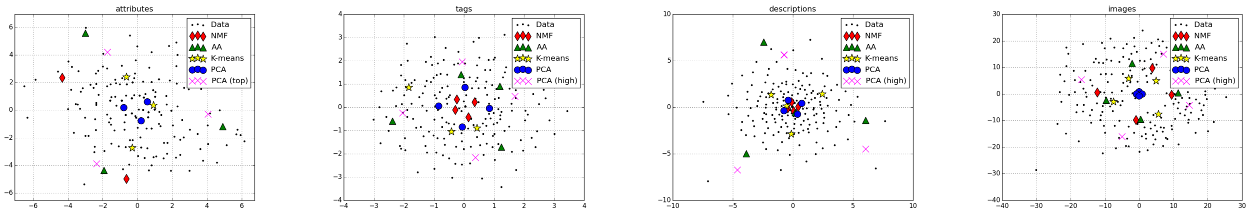
Figure 1 shows the results of analyzing the data set of avatar images. Each row shows the prototype basis image, followed by its top five-weighted individuals (in descending order.)

(a) NMF (b) AA (c) PCA (d) K-means

Figure 1: The figures show the basis images (first column of each sub-image) and their top five top-weighted samples when across the different algorithms for 4 basis components.

Parts-based vs. holistic avatar prototype basis images The prototype images of both AA and KM are holistic, meaning that resultant images can on their own be taken to be complete avatars. This is due to AA prototypes closely matching actual avatar images and K-means prototypes averaging out avatar images. Reconstructing prototype images of PCA do not work (as shown) since they results are eigenvectors, not points in the feature space. Prototype images of NMF appear mixed between holistic and parts-based images. P1 of NMF depicts capes/cloaks while P2 and P3 appear holistic, and P4 bearing a silhouette-like image.

Image consistency of top-weighted individuals Here, we consider the top-weighted individuals for each prototype.
NMF: All individuals associated with P1 have capes, and a majority of them have wings – aspects visible in P1. P2 and P3 appear to depict male and female avatars respectively. All male avatars of P3 have swords, shoulder pads, and gauntlets, and female avatars of P3 have swords, boots (dressed for battle.) This contrasts with female avatars of P4 that feature non battle-oriented attire, exemplified in P4's image of a silhouette with a smaller body frame. Thus, NMF prototypes categorize avatar images into distinct categories, each possessing perceivably similar visual characteristics.
AA: Each prototype image corresponds closely to their top-weighted individual, which is expected for AA. The categories appear distinct from one another like for NMF. For P1–P3, the top-weighted individuals all have visual similarities to their respective prototypes, especially color as seen with P2. P4's images resemble female avatars with numerous accessories (e.g ears, flowers, ponytails, wings.)
PCA: Focusing on the top-weighted individuals for each prototype, it is hard to determine any distinct visual characteristics both between and within prototypes. We note that very few female avatars are seen. Perhaps reducing the data set of images, each of represented as high dimensionality vectors ($32 \times 48 \times 4 = 6144$), into just k=4 principal components was insufficient for representing the variance.
K-means: The images of top-weighted individuals within each prototype do not appear to have much similarity between them, whether in shape or color. However, the prototype images appear distinct from each other – P1 has visi-



(a) Attributes (b) Tags (c) Descriptions (d) Images

Figure 2: The scatter plots visualize how the prototypes obtained from the clustering approaches differ for attributes and text.

bly male avatar characteristics, P3 and P4 have visibly female avatar characteristics, while P2 shows female characters possessing shorter hairstyles.The prototypes are thus aggregated averages of the top-weighted individuals.

Discussion

Figure 2 visualizes the data set and clustering results. For PCA , we show both eigenvectors and top-weighted points.

With a small number of features, as we had for attributes (M=6), we observe that both AA and NMF result in prototypes that lie on the boundary of the data set, compared to those of K-means, which lie closer to the center. As M increases, as for tags (M=397) and text descriptions (M=500), NMF prototypes become located nearer the center, while those of AA and PCA remain on the boundary. This suggests that with small M, NMF's parts-based prototypes will seek to load high on one feature, while minimizing the rest. The small difference between extremal points from AA and NMF suggests that players do not seek to create overly unbalanced characters, likely from our constraint that all 27 attribute points had to be allocated. Finally, with very large M for images (M=6144), all algorithms are located near the center and far from the boundaries. This suggests that (1) k=3 prototypes is insufficient and (2) the number of iterations to convergence is too low. A categorical representation of the set of assets might provide a more structured analysis, but our intention was to gain insight into images in a bottom-up manner, minimizing designer knowledge. Future work will model the discrete categorical and item choices.

We observe that the number of clusters required differs based on both algorithm and data being analyzed. From the scatter plot of descriptions, AA performs well in identifying extremal individuals (archetypes) even with a moderately high number of features M. For images, all algorithms (particularly AA) result in prototypes close to one another, suggesting that a larger k is required. This highlights that greater nuance exists for visual properties of avatars. Another reason could be that restricting $M = 500$ models of text descriptions reduced the complexity of the problem. The plot of tags (not shown) resulted in prototypes that had similar *relative positions between algorithms* as with descriptions, but were located near the center of the data set. The proximity of the prototypes *both within and between* algorithms reflect our results for text, where prototypes either shared exact (or synonymic words) or had similar themes.

Limitations & Future Work These findings reveal how different clustering algorithms work for different technical components. While our choices of the number of prototypes and number of features for our models served our purpose to provide a comparison between approaches, further work into investigating the effects on both algorithmic performance and balancing it with interpretability of prototypes is warranted. We are are developing games based on these player models (particularly AA) to evaluate individual players' perception of cognitive categories (Lim and Harrell 2015b). Evaluating these models for different videogame genres and participant demographics (e.g., self-identified gamers vs. non-gamers) would likely provide alternative findings. We could also consider the use of self-organizing maps (Kohonen 1998; Ultsch 1999), or k-medoids (Kaufman and Rousseeuw 1987), which finds cluster medoids that represent data points (Bauckhage, Drachen, and Sifa 2014).

Conclusions

In this paper, we compared different AI clustering techniques and their effectiveness in modeling players' values through their avatars. We found that NMF, AA, and PCA are effective for finding distinct player notions on character classes and roles in numerical statistical attributes, but only AA and NMF produce prototypes that are directly interpretable and valid. For textual data, NMF, AA, and PCA can identify distinct differences between players, with AA identifying more unusual descriptions compared to NMF and PCA. For images, NMF identifies distinctive parts-based visual characteristics such as the presence of accessories or the relative size of avatars, AA identifies notable characters images that are "best examples," which players strive toward, and K-means form averages of clusters of players mainly by their gender and clothing. Our findings show that such data structures can be effectively analyzed to reveal aspects of players' real-world values and preferences through their customization choices. We believe that combining data-driven clustering models with other methods like qualitative evaluation in a mixed-methods approach has strong implications for developing better computational models of players.

Acknowledgment

We thank the 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.

References

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

Bauckhage, C.; Drachen, A.; and Sifa, R. 2014. Clustering game behavior data. *Computational Intelligence and AI in Games, IEEE Transactions on* PP(99):1–1.

Canossa, A.; Martinez, J. B.; and Togelius, J. 2013. Give Me a Reason to Dig Minecraft and Psychology of Motivation. *In Proceedings of IEEE Conference on Computational Intelligence and Games*.

Cutler, A., and Breiman, L. 1994. Archetypal analysis. *Technometrics* 36(4):338–347.

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 *Computational Intelligence and Games, 2012 IEEE Conference on*, 163–170. IEEE.

Drachen, A.; Thurau, C.; Sifa, R.; and Bauckhage, C. 2014. A comparison of methods for player clustering via behavioral telemetry. *arXiv preprint arXiv:1407.3950*.

Drachen, A.; Canossa, A.; and Yannakakis, G. N. 2009. Player modeling using self-organization in tomb raider: Underworld. In *Computational Intelligence and Games, 2009. CIG 2009. IEEE Symposium on*, 1–8. IEEE.

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

Eugster, M. J. 2011. Archetypal athletes. *arXiv preprint arXiv:1110.1972*.

Harrell, D. F., and Veeragoudar-Harrell, S. 2012. Imagination, computation, and self-expression: Situated character and avatar mediated identity. *Leonardo Electronic Almanac* 17(2).

Harrell, D. F. 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.

Harrell, D. F. 2013. *Phantasmal Media: An Approach to Imagination, Computation, and Expression*. MIT Press.

Jolliffe, I. 2005. *Principal component analysis*. Wiley Online Library.

Kaufman, L., and Rousseeuw, P. 1987. *Clustering by means of medoids*. North-Holland.

Kohonen, T. 1998. The self-organizing map. *Neurocomputing* 21(1):1–6.

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

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

Liberated Pixel Cup. 2015. http://lpc.opengameart.org.

Lim, C.-U., and Harrell, D. F. 2015a. Developing computational models of players' identities and values from videogame avatars. In *Proceedings of the 10th International Conference on the Foundations of Digital Games*.

Lim, C.-U., and Harrell, D. F. 2015b. The marginal: An ai-driven game to computationally model the cognitive perception of category gradience with players' avatars. In *Proceedings of the 2nd AIIDE Workshop on Experimental AI and Games*.

Lim, C.-U., and Harrell, D. F. 2015c. Revealing social identity phenomena in videogames with archetypal analysis. In *Proceedings of the 6th International AISB Symposium on AI and Games*.

Mack Looseleaf Creator. 2015. http://www.geocities.jp/kurororo4/looseleaf/.

MacKay, D. J. C. 2003. *Information theory, inference and learning algorithms*. Cambridge university press.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; and Duchesnay, E. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830.

Rosch, E. 1999. Principles of categorization. *Concepts: Core readings* 189–206.

Seth, S., and Eugster, M. J. A. 2014. Probabilistic archetypal analysis. Technical report, arXiv.org.

Shaker, N., and Abou-Zleikha, M. 2014. Alone we can do so little, together we can do so much: A combinatorial approach for generating game content. In *Tenth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Steele, C. M., and Aronson, J. 1995. Stereotype threat and the intellectual test performance of african americans. *Journal of personality and social psychology* 69(5):797.

Tekofsky, S.; Van Den Herik, J.; Spronck, P.; and Plaat, A. 2013. Psyops: Personality assessment through gaming behavior. In *Proceedings of the Foundations of Digital Games Conference*, 12–65.

Thurau, C.; Kersting, K.; and Bauckhage, C. 2009. Convex non-negative matrix factorization in the wild. In *Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on*, 523–532. IEEE.

Ultsch, A. 1999. Data mining and knowledge discovery with emergent self-organizing feature maps for multivariate time series. *Kohonen maps* 46:33–46.

Yee, N., and Bailenson, J. 2007. The proteus effect: The effect of transformed self-representation on behavior. *Human communication research* 33(3):271–290.

Yee, N.; Ducheneaut, N.; Yao, M.; and Nelson, L. 2011. Do men heal more when in drag?: Conflicting identity cues between user and avatar. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 773–776.