Understanding Video Games: The Quest for an Experiential Fit

Chih-Yu Chiang, Shun Wang, & Hexun Lee
Final Project, Computational Content Analysis

Contents

I. Introduction

Many products today find their way into diverse consumer pools based on their objective quantifications – the horsepower of a car, the specifications of a computer, and the durability of a certain kind of clothing. But it becomes far harder to classify objects when they are unquantifiable, or where some element of subjectivity is involved. Movies, books, and games fall into this category – what we call experience-intensive products. Genres (first-person shooter, role-playing game, puzzle) may be another simplistic way of classifying games; but it is hardly indicative of the rich depth of in-game experiences that one encounters as a player. There is a dearth of research on classifications for experience-intensive products. While players may rely on “expert critics”, who easily judge the worth of such a product based on a game’s playability, we wish to create a kind of classification that is solely based on the particular experiences games invite players to partake in – experiences which are intrinsically neutral and cannot be judged on a scale. Today, *Nintendo* consoles still retain a large fan-following, even if their hardware is not at the forefront of the game industry[1]. This tells us that the types of games and thus, the experiences they invite players to enjoy are far more important than other quantifiable factors such as in-game physics, graphics, and the machine the game runs on.

The social game of interest to us in this project is the experience that players look for in video games. Every game is in itself a contained local landscape that structures the world of the player. Within it, there are social structures, e.g. multiplayer modes, customs and various modes of interaction with non-playable characters and the like; cultural systems, e.g. languages, customization choices; and material resources: in-game currency, the internal environment of the game. In playing games, there is feedback between the game developers and the gamer – and gamers are always on the lookout for experiences the prefer. In other words, good games clearly know their audiences; they know what experiences their players want and know how to replicate and even heighten these experiences. Some developers with a poor consideration of the gamer may fortuitously find a successful formula, but go on to make a sequel that utterly ruins their franchise. In short – they do not understand how experiences can be categorized and marketed to different groups of gamers. We propose that there is a specific experiential fit for each type of video game to individuals, and this can create a better system of product evaluation. In other words, we believe that different players are motivated to seek different in-game experiences.

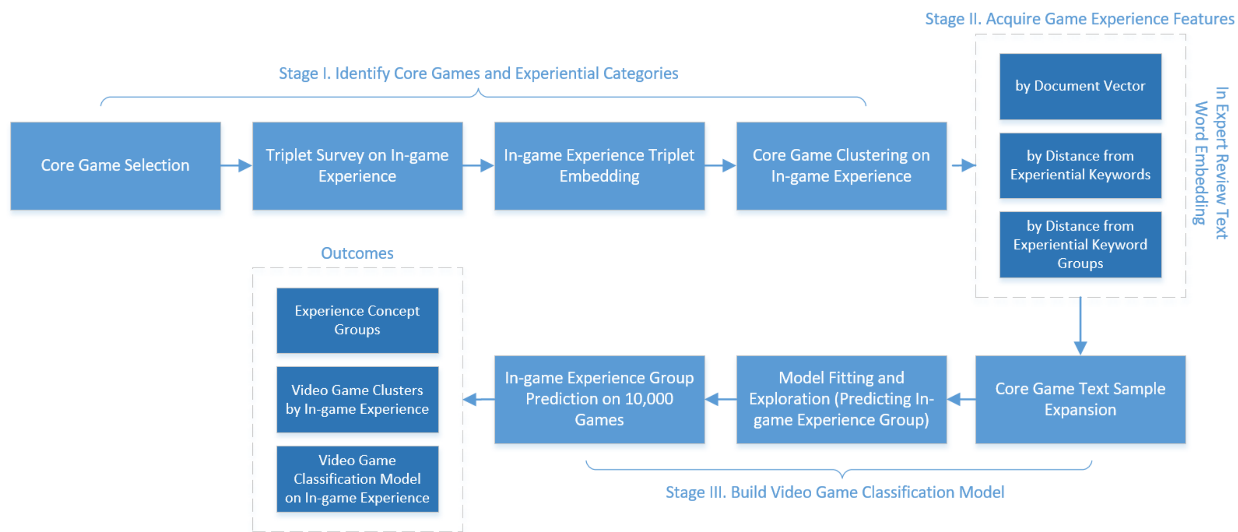
Classifying all the games in the universe is an impossible task. But it seems that if we want to make any substantive claims about gaming experiences, we are expected to have some kind of exposure to the games in question. To circumvent this, we use text analysis of game reviews written by experts to extract the communicative components of these social games; that is, these texts describe for us the different in-game experiences that we will use for classification.

Through this research, we intend to develop a new scheme which describes products by their experiential types, thus injecting a new perspective to the designing of experiential products: one based on the emotions, sensations, and motivations supplied during the products’ consumption experience. When a robust classification system premised upon gaming experiences exists, developers can better target their products to a certain group of players, and players can better evaluate their gaming options. Further, it may also be applied to other experience-based products like movies and even books.

[1] Arthur Gies, “How will the Nintendo Switch compare to the PS4 and Xbox One?” *Polygon*, Published 20 October 2016, Retrieved 12 March 2017 from http://www.polygon.com/2016/10/20/12288138/how-powerful-will-the-nintendo-switch-be-xbox-one-ps4-comparison.

II. Methodology

To categorize video games by their in-game experiences, we implemented a three-stage process—I. conduct a triplet survey to categorize 25 core games, II. tag the games using expert review texts and experiential keywords, and III. build a prediction model classifying video games outside the core games.



A. Identify Core Games and Experiential Categories

1) Core Games Selection

We first wanted to cluster games based on the in-game experiences existing gamers. This required us to find a small subset of games that have had broad appeal over the last few decades. Since we wanted users on Mturk to participate in our survey, we had to select games that were not only varied in genre but also popular enough (also considering the time and financial costs feasible to this project). Across the various genres, then, we selected 25 games that sold an upwards of 5 million titles. The 25 games, hereon known as "core games", are listed below:

Final Fantasy VIII	RollerCoaster Tycoon	Minecraft	Grand Theft Auto V
Counter-Strike	Need for Speed: Underground	Angry Birds	Call of Duty: Black Ops
Tony Hawk's Pro Skater	Mario Party DS	Fifa 16	Wii Play
The Sims III	Nintendogs	Mario Kart Wii	Half-Life 2
Halo 3	Civilization V	New Super Mario Bros	Resident Evil 5

2) Triplet Survey on In-game Experience

Next, we enlisted the help of survey participants on Mturk to determine a good "similarity function" between objects that would help us produce clusters. We wanted to cluster games according to their in-game experiences, but having players describe them one game at a time would be too onerous a task. Instead, we created pair-wise comparison questions that tasked participants to compare in-game experiences and decide which were similar. This helped to reduce fatigue on survey participants, at the same time "alleviating the need to reconcile individuals' scales of similarity."[2] The intuition behind the selection is this: imagine if you are a game store to buy a game for a friend; but the game you have in mind is out of

[2] Omer Tamuz, Ce Liu, Serge Belongie, Ohad Shamir, and Adam Tauman Kalai. "Adaptively learning the crowd kernel." arXiv preprint arXiv:1105.1033 (2011), p. 2

stock. Which other game would you choose based on the similarity of in-game experiences to your original title? Out of the 25 core games, participants selected 5 that they are well-acquainted with and compared them with others. For example, "based on your in-game experiences, is Half-Life 2 more similar to RollerCoaster Tycoon or Metal Gear Solid?" To avoid suboptimal responses, participants must correctly answer factual questions about the 5 games they selected.

Our reasoning for our figures, including selecting 25 core games, 5 familiar games and 20 triplet comparisons for each participant, was based on the research of Tamuz et al. According to this paper, each item should at least show up in 30-40 comparisons in order to get a stable structure. As a result, we designed our Qualtrics survey to have a total of 135 participants, each making 20 comparisons. Filtering out invalid answers, we received 2,690 triplet comparisons. The questionnaire can be accessed here.

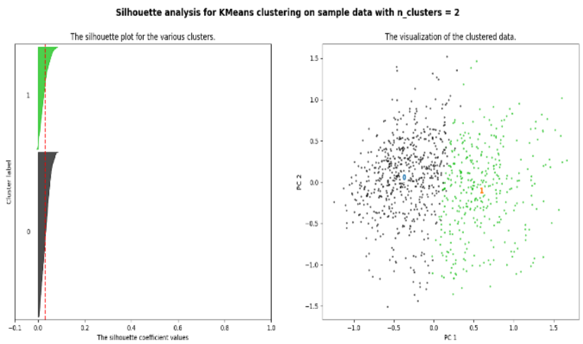
3) In-game Experience Triplet Embedding

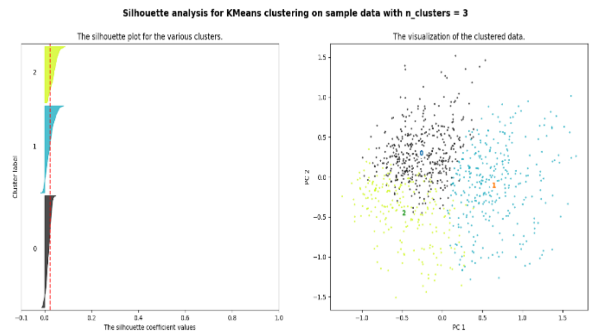
To determine a good "similarity function" between objects – a kernel was built to help with clustering. To do this, we employed t-Distributed Stochastic Triplet Embedding (t-STE) on our survey results. The method, developed by Laurens van der Maaten, specifically adjusted to the task of extracting the information from triplet comparisons. This method generates an embedding that refines the variation between observation points in the original high dimensional space created by the triplet raw data, outperforming GNMDS (Generalized Non-Metric Multidimensional Scaling), CKL (Crowd Kernel Learning), and other existing techniques on the specific triplet data form. The embedding extracted a 25-feature vector for our following analysis.

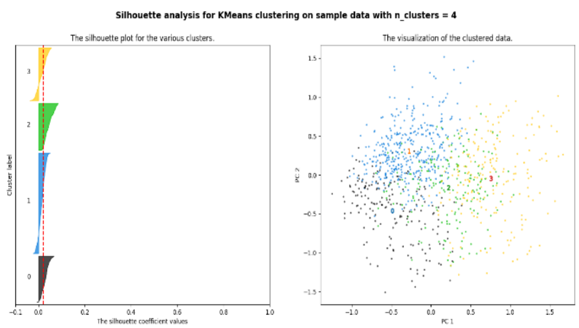
4) Core Games Clustering

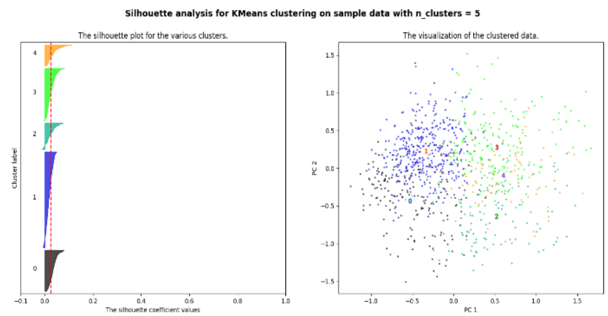
Based on the data we got from the triplet comparison, we conducted K-means as the method to cluster the core games we found. Here we did not use hierarchical clustering since it seems unlikely for a hierarchical structure to be present in video games. Next, Silhouette analysis was applied, with scores from 2 – 7 clusters presented below. Based on the Silhouette scores, and the extent to which we could explain the meaning of each cluster, we chose 5 as the optimal number of clusters for the core games.

Number of clusters	2	3	4	5	6	7
Average silhouette score	0.107151	0.096037	0.099021	0.1036	0.117983	0.133346





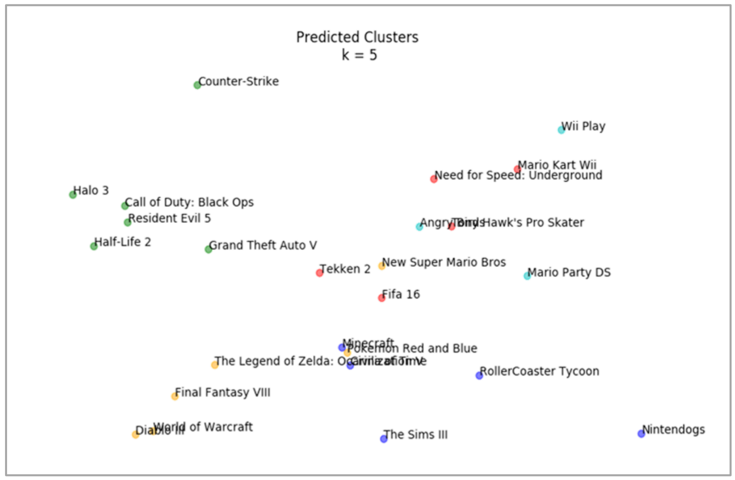




The resulting five clusters are:

Competition	World-based	Violence	Creation/Simulation	Episodic
Tekken 2	The Legend of Zelda: Ocarina of Time	Counter-Strike	Nintendogs	Mario Party DS
FIFA 16	Final Fantasy VIII	Halo 3	The Sims III	Angry Bird
Need for Speed: Underground	Diablo III	Grand Theft Auto V	Minecraft	Wii Play
Mario Kart Wii	World of Warcraft	Call of Duty: Black Ops	RollerCoaster Tycoon	
Tony Hawk's Pro Skater	Pokemon Red and Blue	Resident Evil 5		
	New Super Mario Bros			

We provide tentative labels to the above 5 clusters based on our understanding of their in-game experiences. Cluster 1 we label "Competition". The games in this cluster often (but not always) feature multiplayer modes where players pit their skills against other players or non-playable characters. These games typically require long periods of tension and anticipation before the player hits the "sweet spot" that brings a swift victory. The second cluster we label as "World-based". Every single game within has an elaborate world in which the player, as the protagonist, is fully immersed in. There are intricate maps, currencies, non-playable characters and its very own social structure that would not make sense to one unacquainted to the games. The third cluster we label as "Violence" as every game involves some kind of in-game bloodshed and shooting. They are all unsuitable for younger audiences due to mature content within the game. Cluster four we name "Creation/Simulation". Some games within, like Minecraft and The Sims IIII, are really "sandbox" games that allow players creative freedom to build, fashion, and customize. RollerCoaster Tycoon is a game that simulates the management of a theme park, while Nintendogs simulates the real-life ownership of puppies. Finally, the last cluster appears to us as non-continuous, episodic games. These might be games that one plays once in a while to blow off steam; and importantly, these are games that facilitate quick engagement and termination. In other words, players of Episodic games should find it easy to pick up and put down these games; and these are games that players typically are not addicted to – as in-game experiences are short, relatively finite, and repetitive.

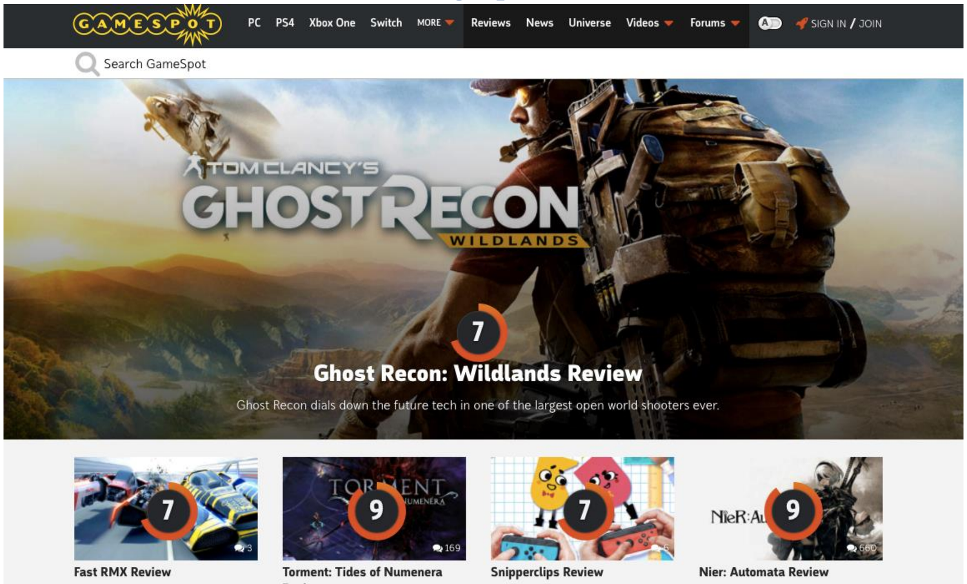


We used PCA projection to chart these games onto a two-dimensional space as shown in the figure on the left. As we can see, there are five different colors that represent the five different clusters. Aside from several overlaps, the clusters are relatively distinct. Interestingly, Tekken 2, from the Competition cluster, is closest to the Violence clusters, which corresponds to its in-game experience as a fighting game. Also, the most dispersed cluster is the Episodic cluster, which contains Wii Play, Angry Birds, and Mario Party DS. This is also understandable, because their actual in-game content are different from one another - one is a puzzle game, another is a multiplayer sports game, and the last, Mario Party DS, is a collection of multiplayer minigames. If there is any conclusion to be made, it is that many games have in-game experiences that are not exclusive to their main cluster label. In other words, many games do not sit neatly into a particular cluster, and games do in fact share similarities across clusters. Expanding the number of clusters would provide a more ideal grouping for the games. However, given the limited number of core games we were testing, we avoided spreading the clusters too thin.

B. Acquire Game Experience Features

From the reviews we collected, we tried to find features for the clusters we got from triplet comparisons to predict new game clusters based on their review articles. Three further approaches were utilized to obtain features: 1) document vectors of game review articles in the game review word embedding space as the features, 2) distance between the document vectors and experiential keywords as the features, and 3) distance between the document vectors and experiential keyword groups as the features.

1) Expert Review Texts Word Embedding Space



More than evaluating the value of a game, reviews help to convey in-game experiences to its readers. They typically relay the various experiential dimensions within a game: kinds of in-game systems the game has, the various challenges players will encounter, storylines, and the various intricacies within the game's social world. Since we cannot possibly play every existing game in the universe, we relied on expert reviews' document vectors in a word embedding space to provide features as the first step for building our predictive model. In applying document vectors, each representing the centroid of an entire article, attempts were also made to preserve as much of the reviews' most essential content as possible.

11,022 game reviews covering 9,805 unique games from Gamespot were scraped. Launched in 1996, this major video gaming site in the U.S. received more than 60 million visitors in 2008. Gamespot provides independent expert game reviews, it is a pure information provider (through blogs, news, and review articles) uninvolved in the video game business. Through our 11,022 game reviews, we amassed a total of 16,643,915 words or analysis. The length of reviews averaged 1,170 words – meaning that on average, 1,170 words capture the essence of in-game experiences in a game. Stopwords were removed using the NLTK English stopword dictionary; no stemmers and lemmatizers were used as we wanted to preserve the information from the sentence structures. We performed sentence-based word embedding, acquiring 300 features.

2) Experiential Keywords

However, not all content in the reviews are amenable to our goal of classifying by experiences. That is, not all descriptions within reviews concern experiences—many are evaluations of a game's value, such as the price of a game or the production quality of a game. These factors could negatively influence the accuracy of the model, and affect the prediction outcome. As a result, we tried to parse through the content using experiential keywords extracted from the dictionary as our second approach in extracting features for our model. We identified the "experiential keywords" and calculated the distance between the game expert reviews and each of the experiential keywords as the features.

To do so, we first searched through Webster's Unabridged dictionary's 105,000 word entries (1913 edition, updated by some transcribers, released and licensed from Project Gutenberg) to identify words whose definitions include the seed descriptions. We tested a variety of words as the seeds:

- Words describing surroundings, environment: Semblance, Mood, Tone, Feel, Impression, Taste, Flavor, Sense, Spirit, Psyche, Inclination, Tenor, Milieu, Sensation, Ambience, Emotion.
- Words that provoke responses: Encourage, Motivate, Stimulate, Excite, Actuate, Energize, Impel, Incite, Innervate, Inspire, Lead, Encounter, Move, Propel, Experience, Rouse, Tap, Incentivize.
- Words describing appealing experiences: Immerse, Encapsulate, Concern, Consume, Engross, Preoccupy, Rivet, Obsess, Fascinate, Fill, Involve.

Filtering away irrelevant results, Emotion, Feel, Experience, Encounter, and Sensation were the 5 seeds we used to generate our experiential keyword list as their results were the most fruitful. The search provided us with 1,168 words, which we trimmed to 1,166 after cleaning the data. Further filtering out the keywords not used in the video game expert review articles, the final experiential keyword set including 606 words, which provided us 606 features for our model via their cosine similarity against the review articles. Some randomly selected sample keywords are as follow:

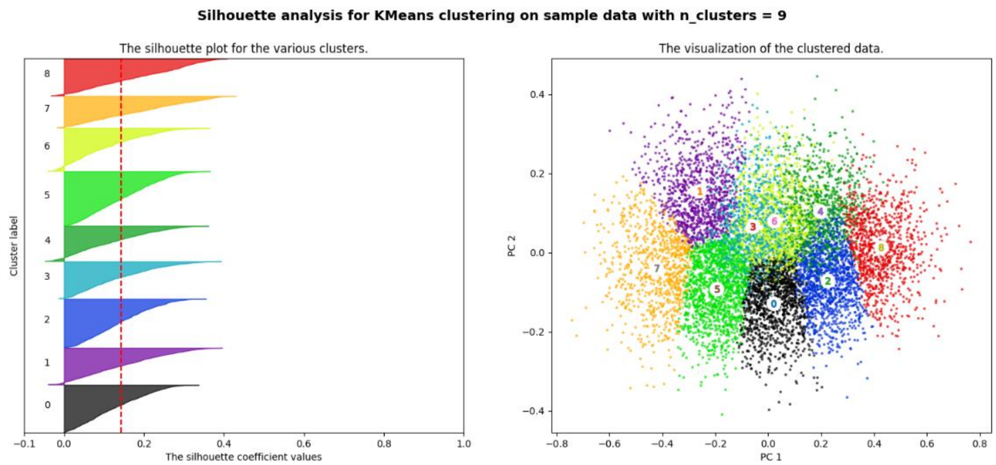
ANIMAL	BILE	INTERJECTION	INSOLENT	NERVOUS	FUMBLE
TEMPERATURE	APATHETIC	EQUIVOCAL	DISDAIN	SCANDALOUS	FRONT
LOATHFUL	APATHETICAL	PRACTICALLY	INSTINCTIVE	DECLAMATION	AFFECTUOUS
DESIGN	NUMBNESS	RIVAL	ACUTE	USAGE	QUALM
ALIVE	FERVENT	REASONING	ACTION	IMPETUOUS	DELICIOUS

3) Experiential Keyword Groups

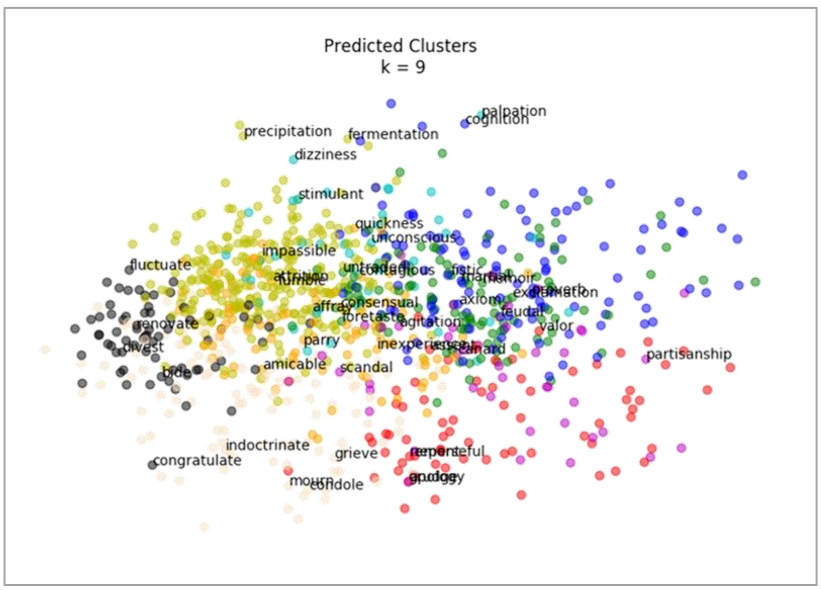
With more than 1,000 experiential keywords, we tried to further categorize them into groups as if to sieve out the noise brought in by the large number of features. With too many potentially irrelevant features, the model could easily overfit training data and miss the pattern of interest to us, damaging its efficacy. To overcome this, we used a Google News word2vec embedding, pretrained by 100 billion words from a Google News dataset with 300 features, to locate the 950 word vectors. We applied a word embedding space different from the one from game review articles so that we could acquire a more general and precise relationship between the experiential words—the Google News word embedding represents a more generalized usage of these keywords. Compared to the Google News word embedding, the game review word embeddings contained more video game specific semantics, which could be misleading due to their particularities (especially their cultural usages) when categorizing these experiential keywords.

Next, experiential keywords were clustered with Ward and K-means methods. Our team used a hierarchical approach to conduct the keywords clustering because there could plausibly be a hierarchical structure for the experience-related words. For instance, words that relate to "positive experience" could contain many subtypes including exciting (high arousal level), curiosity (exploring and openness) and so on. Unfortunately, we did not find such pattern in the hierarchical cluster; instead we found that the cluster could be organized around different types of words, such as nouns, verbs and adjectives, which makes the categorization unfeasible for our purposes.

Next, the k-means method was attempted. Based on the Silhouette analysis, we chose 9 as the optimal number for the clusters. It should be noted that the average silhouette scores for 5 and 6 clusters were larger than 9 clusters; and our team found that these clusters were too general to explain specific types of experiences. Thus, we choose 9 as our final number of keywords clusters.



Number of clusters	5	6	7	8	9	10	11
Average silhouette score	0.0244385	0.0213564	0.0147123	0.0061013	0.0184427	0.0142095	0.0110546



Finally, we computed the centroid of each experiential keyword group in the game expert review word embedding space and calculated cosine similarities between review document vectors and each of the 9 keyword group centroids. These results were then taken to be the features to be used in the following model.

C. Build Video Game Classification Model on In-game Experience

With 1) the core games and their in-game experience labels and 2) the various features we extract from the video game expert review articles, our team built a model predicting the in-game experience labels and applied the model to video games outside the core game set.

1) Core Game Text Sample Expansion

Since we only had 25 core games in the survey, to make the model more robust, the review articles database for these 25 core video games was expanded to include other game review websites, such as IGN, Gamefaqs, and PCgamer. The corpus of core game reviews then, contained a total of 206 review articles. Each review article makes up one individual observation (game) in our model building process.

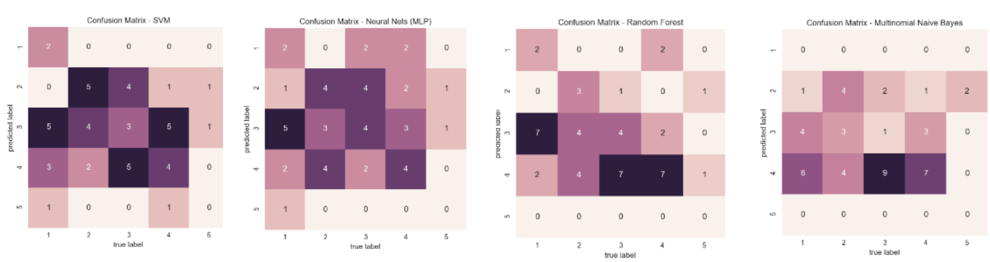
2) Exploration and Fit

With the 206 observations, our team tested the 3 sets of features in the game review embedding space, the document vectors (300 features), the distances between document vectors and each keyword (around 600 features), and the distances between doc vectors and each keyword group (9 features), each with 4 different models: SVM, SVM, Neural Nets (Multi-layer Perceptron (MLP)), Random Forest and Naive Bayes (multinomial). To test these models' accuracy, we did a simple cross-validation with the models trained on 70% (144 observations) of the core games' reviews and tested on the remaining 30% (62 observations) and computed the coefficients and the confusion matrices. The most important coefficient for us was the F1 score, which can be interpreted as a weighted average of the precision and recall.

<u>Document vectors (300 features)</u>

First, we built a model based on the game reviews' document vector in the game review word embedding space. Results of the model's prediction on the testing set are as follows:

SVM | Neural Network | Random Forest | Naïve Bayes

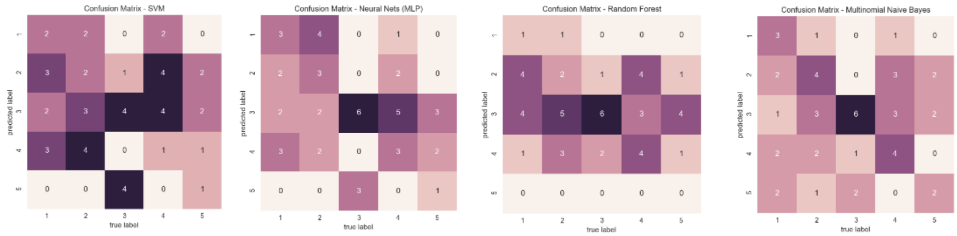


	Precision	Recall	F1
SVM	0.449848	0.297872	0.304354
Neural Network	0.297872	0.297872	0.290829
Random Forest	0.395536	0.340426	0.323004
Naïve Bayes	0.179839	0.255319	0.199917

<u>Distances between document vectors and each keyword (around 600 features)</u>

Second, we built the model on cosine similarities between each experiential word and the game review document vectors. Results of the model's prediction on the testing set are as follows:

SVM | Neural Network | Random Forest | Naïve Bayes

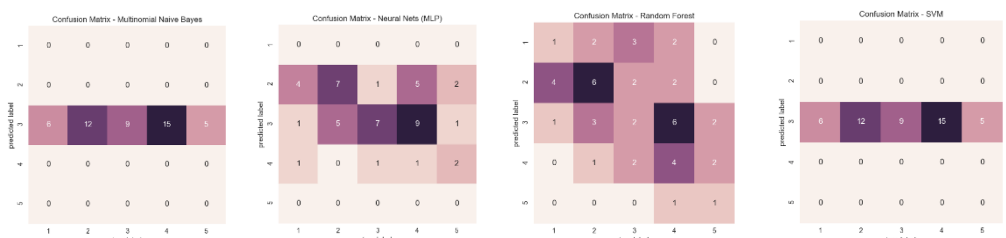


	Precision	Recall	F1
SVM	0.21253	0.212766	0.204339
Neural Network	0.346049	0.340426	0.326444
Random Forest	0.282721	0.276596	0.235395
Naïve Bayes	0.429855	0.404255	0.398854

<u>Distances between doc vectors and each keyword group (9 features)</u>

Lastly, we built the model on cosine similarities between each experiential group and the game review document vectors. Results of the model's prediction on the testing set are as follows:

SVM | Neural Network | Random Forest | Naïve Bayes



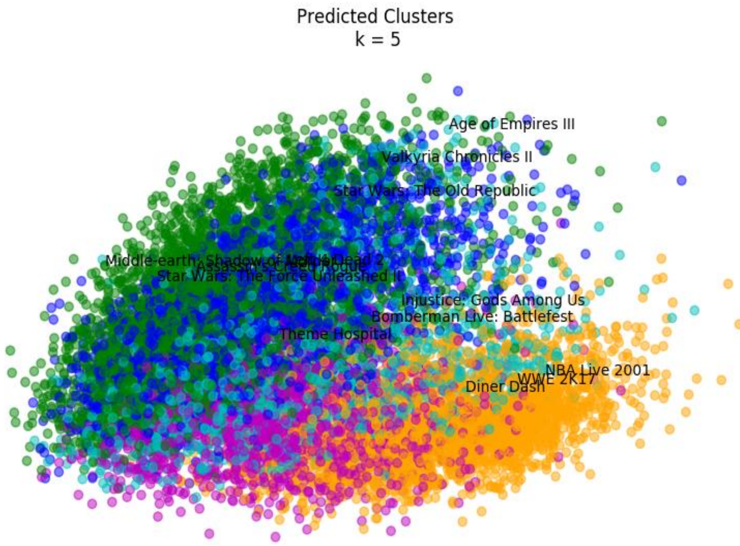
	Precision	Recall	F1
SVM	0.036668	0.191489	0.06155
Neural Network	0.216174	0.319149	0.230997
Random Forest	0.347771	0.297872	0.306157
Naïve Bayes	0.036668	0.191489	0.061550

According to the coefficients and the confusion matrices, the model built using the second approach with Naïve Bayes yielded the most accurate results. While we suspect a potential overfitting problem in using all 600 features in the model, it was still used to build our predictive model for the classification of the remaining reviews.

One reason affecting the precision of the first method might be due to present confounding factors in the review articles. Also, for the third approach, we posit that it might be due to the dearth of features that led to its inability to capture sufficient information.

3) Application on 10,000 Games

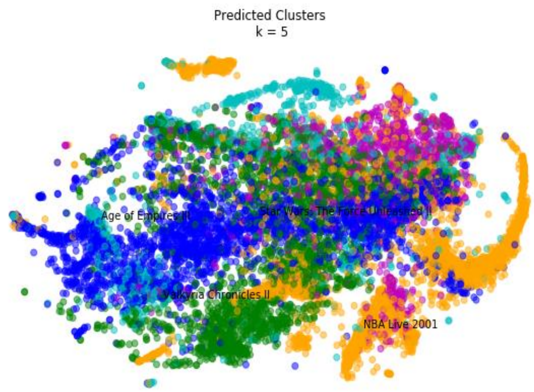
With the Naïve Bayes model, the model was applied to more than 10,000 other video game reviews we collected from GameSpot, with more than 90% of them being reviews of a unique video game. This effectively gave us more than 9,000 predictions of in-game experience cluster labels of games outside our 25 core game set. In the figure below, we project the result via a PCA dimensional reduction:



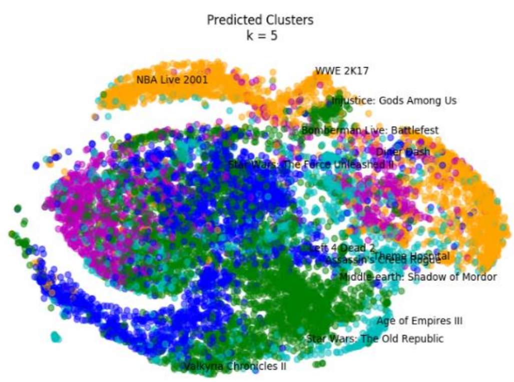
Experience label predictions with PCA projection
—using distance between the document vectors and experiential keywords (600 features)

The results were promising. According to the figure above**,** Diner dash, WWE 2K17 and NBA Live 2001 were clustered together – these are games that emphasize speed and physicality. Notice also that the two Star Wars games belong to different clusters. One possible explanation is that Star Wars: The Old Republic is an online roleplaying game, and central to its gaming experience is player-vs-player battles, player-to-player interactions, and collaborations with players for purposes such as defeating in-game bosses etc. Within this cluster also, is Age of Empires III, which has a popular online multiplayer mode.

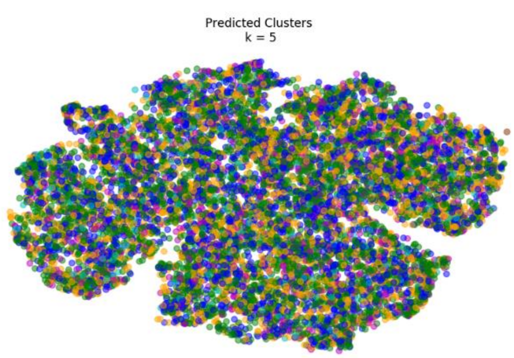
With the tSNE projections of roughly 10,000 testing points and their labels as shown in the figures below, we can gain some superficial insights to the quality of our model. Compared to the 600-feature model, the one with 300-features performed slightly worse as the pink, green, and light blue clusters are basically indistinguishable from each other. The projection with the 9-feature model did not produce discernible patterns confirming our opinion that it is an inferior model through cross validation with our 25 core game samples. Based on the figures, the model which used the 600 experiential keywords as its features, clearly showed an improvement in clustering than the others, better capturing the experiential differences of the new games.



Experience label predictions with tSNE projection
—using document vectors of game review articles (300 features)



Experience label predictions with tSNE projection
—using distance between the document vectors and experiential keywords (600 features)



Experience label predictions with tSNE projection
—using distance between the document vectors and experiential keyword groups (9 features)

III. Outcomes

We had three major outcomes. First, we produced 9 experience concept groups obtained from the 950 experiential keywords; second, 5 video game clusters based on in-game experiences; and finally the development of a classification model able to predict the various clusters for remaining games in the universe.

A. Nine Experience Concept Groups

The figure below reveals the nine keyword clusters we derived from features of our 1,168 experience-based words.

Cluster	0	1	2	3	6	7
Explanation	**Metaphysical**	**Ego**	**Interaction**	**Sensation**	**Disequilibrium**	**Intensity**
Keywords Samples	devoutly	instinctive	taint	numbness	deplore	aridity
	superstition	mood	please	alarm	rhetoric	sentient
	mysticism	impulse	brighten	tissue	revolt	sensationalist
	ethical	loneliness	presage	unconscious	umbrage	peckish
	tribalism	ego	bide	sensory	allegiance	percussion
	philosopher	sensitivity	injure	fever	apoplexy	hum
	nativism	undercurrent	brutalize	temperature	repulsion	delicacy
	canard	consciousness	gird	pricking	penance	bile
	lore	magnetism	teem	contagious	polarization	tactile
	emulation	solitude	deaden	ferment	contrition	titillation
	axiom	virtuosity	sweeten	explosion	sensationalism	spectral
	inequality	kindness	bicker	stimulant	apology	matron
	worthiness	affluence	rue	agitation	partisanship	syllable
	telepathy	smile	chafe	collision	grudge	memoirs
	experimentally	hunger	congratulate	palpation	assent	exclamation
	subjective	sensuality	avenge	dizziness	repent	penitent
	feudal	inexperience	divest	frost	scandal	tentacle
	cognition	greenness	indoctrinate	fermentation	mourn	foretaste
	martial	valor	fluctuate	tartar	grieve	fistic
	proverb	quickness	parry	fumble	condole	memoir

Out of the 9 clusters, we selected six that made sense to us. Every cluster has a unifying label. For example, we label cluster 0 as "Metaphysical" because the words within typically deal with experiences that are outside of the spatial-temporal world. Words like "superstition", "mysticism", "cognition", "telepathy" and "worthiness" are all related categories that are less tangible. Cluster 1 we label as "Ego" as they seem to all relate to the self, or a sense of self – one's mood, impulses, instincts, and so on. Cluster 2 we label as "Interaction" as they typically involve more than one party. To "congratulate" usually is a conduct between parties (other than self-congratulation); and to "avenge", "divest", "party", "indoctrinate", "please", and "injure" all relate to conduct between agents. Cluster 3, labelled "Sensation", relates to experiences in responses to external stimuli: "Pricking", "numbness", "dizziness", "palpitation" are different sensations that are felt. Cluster 6 is labelled "Disequilibrium". The words within this cluster tend to signify friction and imbalance in a previously harmonious relationship. Words like "scandal", "mourn", "grudge", "apology", "polarization" signal an alteration in a relationship between two or more people due to conflict or other circumstances. Finally, we label cluster 7 as "Intensity" as most words serve to elaborate upon the category of "Sensation". The word "aridity" signifies the degree of dryness, while "exclamation" adds emphasis to a claim.

The various clusters of experiential words were sufficient enough for our purposes. We believe that it does provide a rudimentary coverage of the myriad of experiences in human living. Still, we recognize that a theory-based classification of various experiences would be more ideal.

Five Video Game Clusters by In-Game Experience

From the pair-wise survey and triplet embedding, we acquired the five video game clusters by the in-game experiences—Competition, World-based, Violence, Creation/Simulation, and Episodic; we believe they are essential experience types that echo traditional game experience theories.

Competition	World-based	Violence	Creation/Simulation	Episodic
Tekken 2	The Legend of Zelda: Ocarina of Time	Counter-Strike	Nintendogs	Mario Party DS
FIFA 16	Final Fantasy VIII	Halo 3	The Sims III	Angry Bird
Need for Speed: Underground	Diablo III	Grand Theft Auto V	Minecraft	Wii Play
Mario Kart Wii	World of Warcraft	Call of Duty: Black Ops	RollerCoaster Tycoon	
Tony Hawk's Pro Skater	Pokemon Red and Blue	Resident Evil 5		
	New Super Mario Bros			

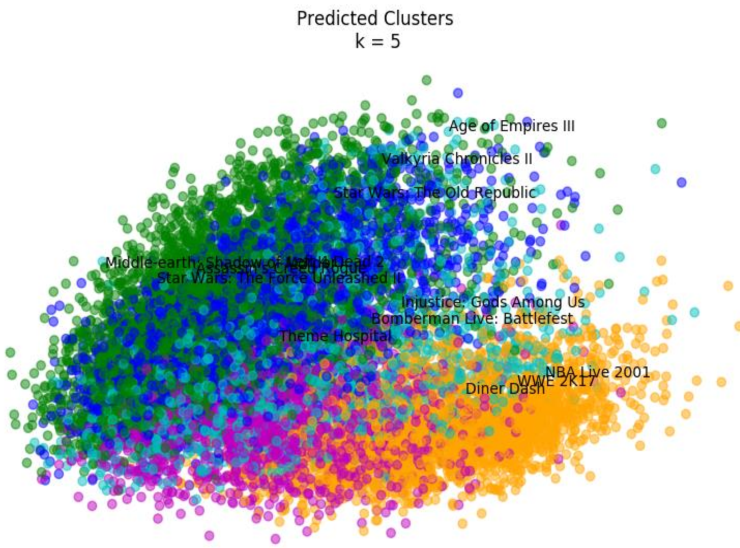
Earlier on, we provided superficial explanations of the clusters that were derived from our personal knowledge of the core games at hand. However, it is interesting to note that upon further examine, our clusters enjoyed a theoretical fit, especially with Bartle's work. Bartle theorized of four classifications for games: Explorers, Killers, Achievers, and Socializers[3]. At first sight, the clusters of our core games – 2, 3, 4 and 5 in figure above appear to correspond to Bartle’s classes Explorers, Killers, Achievers, and Socializers respectively, tentatively supporting Bartle’s classification. Already, our clustering produces a different grouping than traditional methods based on genre would. For example, genre classification would label Cluster 3 as First-Person Shooters (FPS). But notice not all games in Cluster 3 are FPS. For example, Grand Theft Auto V and Resident Evil are not FPS like Counter-Strike or Call of Duty: Black Ops. Rather, if weapons are used, these games are third-person shooters; and one follows a storyline while the other is an open world. Still, it corroborates Bartle’s classification of “Killers”, and all games in Cluster 3 games can be considered “shooter” games, as shooting and killing are main features of the game.

But interestingly, Cluster 1 does not sit neatly into any one of the four categories, offering potential for theory building. More specifically, Cluster 1 games are traditionally classified as: fighting (Tekken 2), sport simulation (Fifa 16; Tony Hawk’s Pro Skater), and racing (Need for Speed: Underground; Mario Kart Wii), and will not typically be grouped into one. We posit that the most intuitive explanation that unifies these games can only be derived through examining their individual gaming experiences. A common thread that all games in Cluster 1 share would be that they all involve experiences of: physicality, thrill, competition, and nervous tension. As a result, players interested in these experiences can expect that, as players, they will be on tenterhooks as winning or losing falls on a split second of attempting to: deliver the critical blow, score the winning goal, edge an opponent in a race, or to pull off a gravity-defying skateboard trick. The suspense involved in attempting something as the clock winds down then, can be the distinguishing experience for Cluster 1, which could be used for theory formation.

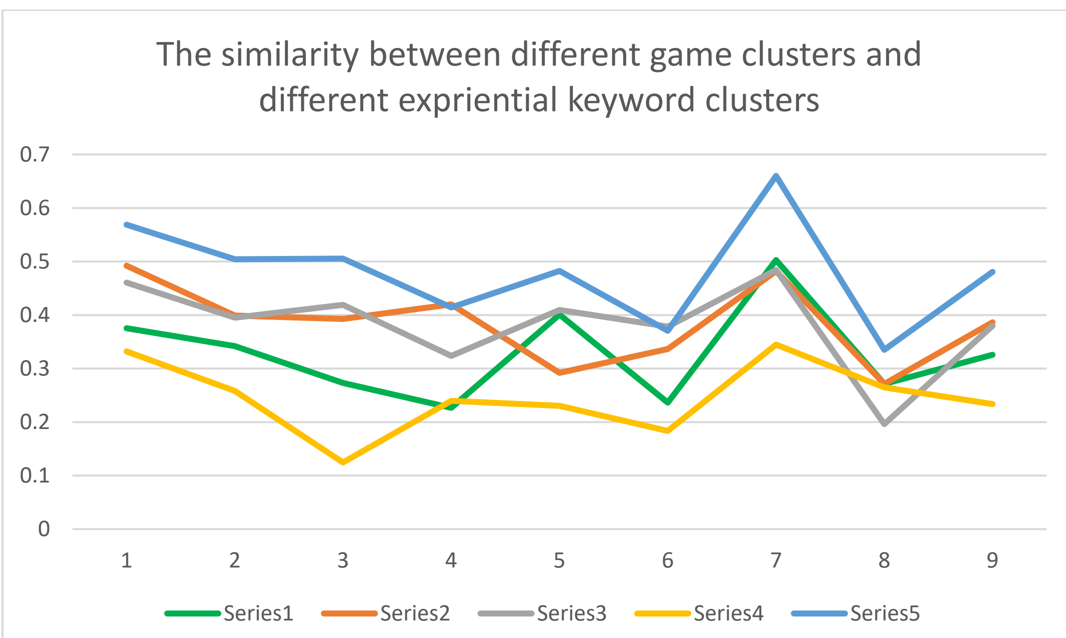
[3] Bartle, Richard. "Hearts, clubs, diamonds, spades: Players who suit MUDs." Journal of MUD research 1, no. 1 (1996): 19.

This new categorization sheds light on our understanding of the root motivations making players involving in video games. Crucially, this categorization helps raise player awareness of their video game preferences and playing behaviors, which could improve their game selection processes and their experience with the video games. Further, it also provides a novel perspective through which video game designers can create better games for their targeted group of players.

B. Video Game Classification Model on In-game Experience



To better describe the clusters, we computed the distances between each game experience cluster and each experiential keyword group and represented the scores in the graph below.



In this graph, the World-based cluster (orange) scored highest on the experiential word cluster of "Sensation". This fits the typical conception of open-world games as they try to create an entire universe with its own unique visuals. As a result, conveying different sights and sounds and sensations are critical to delivering a sense of realism to players. We can take the recently released and highly anticipated Legend of Zelda: Breath of the Wild. Known for its expansive internal world, players of Legend of Zelda are immediately transported into an immersive world replete with breathtaking images and complete with a full treatment of orchestral music, which subtle changes with the situation. In fact, one could even "feel" the gentle breeze by the subtle movements of the grass! There is no doubt that game developers of open-world games know to work with various sensations – primarily audio and visuals to convey such a sense of realism.

There is also little surprise that the Episodic cluster (blue), Competition cluster (green), and Violence cluster (grey) all highly correlate with the disequilibrium word cluster. After all, these three clusters do sometimes involve some kind of domination over others: when someone wins, another has to lose. Such a zero-sum dynamic may explain why games like these can be emotionally-charged and even lead to some relational friction. On the other hand, this zero-sum dynamic is absent in the Creation/Simulation cluster (yellow). According to the graph, out of the 5 game clusters, Creation/Simulation had lowest Disequilibrium scores as an experiential description. Again, it makes intuitive sense: creation/simulation games are usually single-player modes that center around fashioning some ideal image or simulating the management of a project. Such games rarely, if ever, pose defeating challenges that may cause emotional turmoil, but usually try to bring out the creative best in individuals. By and large, then, we do find that our game clusters broadly correspond to our experiential word clusters.

Pure frequency of each experiential keyword was also employed to identify the top words for each cluster of games. For cluster 1, the "Competition" cluster, the wordcloud contains experiential words like “hot”, “enjoyable”, “offensive”, which appear to corroborate with our findings. Cluster 2, the “Open World” group, words like “adventures”, “strange”, “finding”, “worlds” appear – again cohering with the kind of games in the cluster, most of which emphasize the exploratory nature of in-game experiences.

Cluster 3 - Killer

Cluster 4 - Achievement

Cluster 5 - Socializer

The "Violent" cluster produced words such as, “fighting”, “attacks”, “assault”, “shooters”; and cluster 4, the labelled "Creation/Simulation", popular words were "Particular", emphasizing customization, and “career”, which seem to hint at the management and simulation nature of the games. Finally, for the last cluster, the "Episodic" games, “friends” and “party” surfaced.

Admittedly, many other irrelevant or unexplained words still popped up in the wordcloud. However, we find our results encouraging, thus providing us with good reasons to believe that by constructing a more robust model, we will produce even more precise results.

To reiterate, our ultimate objective is to develop a model that could predict the cluster locations of the other games in the universe. This prediction model, after iterative training to make it more precise and robust, can serve as a basis of game recommendation systems to the end video game players and a database where game designers draw inspirations and reference when developing new games.

IV. Limitations and Extensions

Like every other study, our research is not without its limitations. The most obvious one is the difficulty of capturing the essence of in-game experiences in words. The descriptive abilities of expert critics vary, and even the best descriptions of an in-game experience cannot represent the intricacies of an experience that heavily relies on audio-visual content. We are well aware that in the recent years, online streaming websites like YouTube have become popular sites for gamers to seek out game reviews. Admittedly, their ability to represent the very same audiovisuals to audiences is superior to written game reviews. However, there is still value in written game reviews, as they help to verbalize essential in-game experiences for us. A possible extension would be to combine our project with audiovisual representations of in-game experiences for more robust results. For example, we could ask survey participants to compare experiences after perusing game review videos on the web.

Additionally, expanding our project would make our model much more robust. For example, due to the costs of survey taking, we had to rely on our personal knowledge of video games to determine which 25 out of more than a hundred most popular games should be used for the survey. Expanding our core set beyond 25 would also provide us with useful data for cluster building. We also believe that the clusters, and hence the experiences, are not limited to 5 types, and expanding our core set would also enable us to expand the number of clusters for our games. This would allow us to make more precise differentiations between types for in-game experiences. Earlier on, we saw that even though Resident Evil 5 and Grand Theft Auto V were not FPS games, they were clustered with other FPS games. However, Resident Evil 5 and Grand Theft Auto V are still shooters, or at the very least, were games that involved killing, bloodshed, and violence. We predict that an enlarged cluster would enable us to make even more fine-grained differentiations within the shooter cluster itself.

Moreover, we hope that the relationship between personality and in-game experiential preferences could be explored. We did in fact attempt to investigate the relationship between the two by adding a personality scale and a game preference question in the survey. Specifically, we asked participants to evaluate the extent to which they chose games based on the respective in-game experiences. Further, encouraged by previous research that established a strong relationship between people's evaluation of video games and their basic needs in self-determination theory[4], we chose the Basic Psychological Need Satisfaction and Frustration Scale[5] to sort our survey participants into various personality categories. We find that some of our results do support the aforementioned theory. For instance, individuals with high levels of competence satisfaction tend to hold a more positive attitude towards the "Competition games", and persons with a high level of relatedness satisfaction are more likely to be fond of world-based games. We certainly aspire to make more authoritative articulations about the relationship between personality and game experience types in the near future.

With regards to the experience-based word clusters, there are certainly more than 9 experience concepts. Much more research needs to be conducted to add in more keywords that relate to experience. We began to collect the number of keywords relating to experience based on five seed-words: emotion, feel, experience, encounter, sensation. These were simply words that came to our minds in trying to describe elements of a process of "experiencing". In other words, this was a rather arbitrary process, and we believe that it can be improved if we used a more systematic procedure that could churn out relevant experience-based

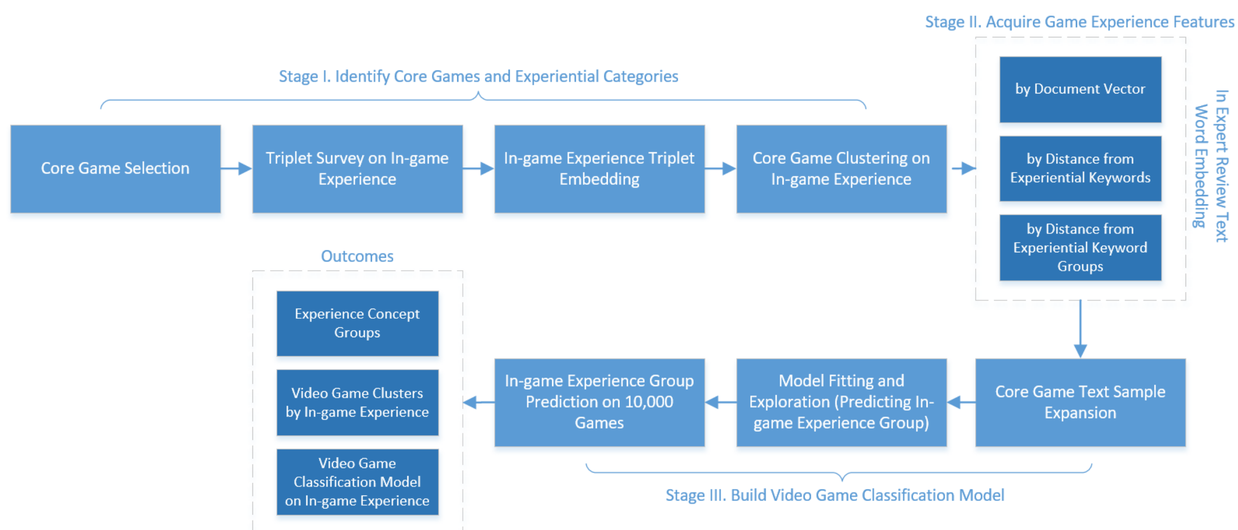
[4] Ryan, Richard M., C. Scott Rigby, and Andrew Przybylski. "The motivational pull of video games: A self-determination theory approach." Motivation and emotion 30, no. 4 (2006): 344-360.

[5] Chen, Beiwen, Maarten Vansteenkiste, Wim Beyers, Liesbet Boone, Edward L. Deci, Jolene Van der Kaap-Deeder, Bart Duriez et al. "Basic psychological need satisfaction, need frustration, and need strength across four cultures." Motivation and Emotion 39, no. 2 (2015): 216-236.

vocabulary. As with more quality vocabularies, we can create more meaningful experiential groups to better capture the "experience space", sieving out the noise while retain the predicting capability.

Next, the basis of our predictive model fundamentally hinges on game reviews, which average 1,170 words. Expanding the number of review articles per game would provide us with more data for analysis. Further, it is unlikely that a single review from one person's perspective would be sufficient to describe all the experiential elements within a single game. We would expect the model's prediction ability be improved by including more than one article per game as more reviews coming from different sources covering different features of a game and thus more complete in-game experience; another good source of texts describing in-game experience in detail is video game walkthroughs, which meticulously illustrate each system and mechanism of a game; the walkthrough texts are recommended to be included in the future as well.

Crucially, one viable extension to our project would be to combine personality and motivation to our classification system. Our existing classification system may be helpful for users with years of gaming experience. In other words, they already have an existing input of games that they enjoy and other games that they enjoy less. If we could find correlations between experiential types and personality types, this would help players with little or no gaming experience to predict the kind of games they are likely to enjoy.



Lastly, we are confident that our results can be generalized and applied to other experiential products. Our methods of identifying experiential keywords, doing triplet surveys and embedding, and tagging the game items by their review texts can be easily be replicated with other consumer products that also are heavily experience-based. Movies serve as an excellent example of an experience-based product. With more and more filmmakers using 3D and even 4D technology, there is no doubt that movies, like video games, are at the forefront of challenging the frontiers of experience generation. Movies are also widely reviewed and discussed in the news, and through reviews and blog articles. This serves both consumers and developers well, as producers can tailor their games/movies to specific audiences, and audiences can easily know their game/movie preferences, thus making more optimal choices.