The Impact of Virtual Reality on Gaming Experience: A Perspective from Half-Life on Steam

Fangyu Yu[1], *Ba-Hung Nguyen*[2], *Duy-Tai Dinh*[3], *Xueyou Wu*[1], *Yang Yu*[1]
Tsutomu Fujinami[1]

[1]Japan Advanced Institute of Science and Technology, Nomi, Japan

[2]Thai Binh Duong University, Nha Trang, Vietnam

[3]The Kyoto College of Graduate Studies for Informatics, Kyoto, Japan

ABSTRACT

As a new form of human-computer interaction, Virtual Reality (VR) introduces new opportunities for game development. Nevertheless, research conducted on the influence of VR on player reviews data was limited. This paper analyzes the influence of VR on the player community of the Half-Life series, including Half-Life 2 (non-VR edition) and Half-Life: Alyx (VR edition) by topic modeling. Particularly, a dataset of approximately 100,000 reviews was first collected from Steam. Then we used the topic modeling, and three experts were involved in analyzing these topics to categorize reviews into 11 topics of Half-Life: Alyx. We also detected the pivotal insight that most game players are concerned with only five topics of fundamental game elements. Furthermore, players' positive ratings were compared on these five topics, and the p-value of the two-sample t-test reveals VR has no significant impact on these fundamental game elements.

Keywords: Virtual Reality, Online Reviews, Topic Modeling, Steam

INTRODUCTION

Virtual Reality (VR) is a technology that produces a digital environment and simulates experience to allow users to interact with this media, giving a perception of being mentally immersed in the environment (Alves et al. 2020). VR has been applied in many applications such as medical education, urban design, and entertainment. VR has great potential in the video game industry. A VR game is a video game played on VR hardware. Most VR games are realized with a head-mounted display unit or headset along with one or more controllers. VR was employed in the video game industry around the 1980s. The size of VR market has grown rapidly. Many game developers have started to post existing games to VR environments as well as uploaded VR games on game platforms such as Steam.

Steam is a video game digital distribution platform maintained by Valve Corporation. Steam also allows players to post reviews of games to indicate whether to recommend to others. Players are prompted to provide their overall feeling regarding the game: "Recommended" - positive review or "Not Recommended" - negative review. The steam review system allows players to easily share game experiences with others and describe how the game meets their expectations. Half-Life is a first-person shooter (FPS) game released by Valve in 1998. Half-Life contains many gameplay interactions common across other popular FPS games. Its successor, Half-Life 2, was released in 2004, followed by its two downloadable contents: Half-Life: Episode One (2006), Episode Two (2007). As the first VR game of this series, Half-Life: Alyx (2020) received praise for its graphics, voice acting, story, atmosphere, and is considered the killer application for VR games.

Although there are many pieces of research focusing on gaming experience on VR (Shelstad et al. 2017, Porter III et al. 2018), little research has been conducted on reviews of VR game players', especially the comparison between VR and non-VR game reviews. Thus, how to effectively detect and unify the aspect-based topics of user experience in different games is still an extremely challenging task (Lin et al. 2019). By comparing the reviews of the VR and non-VR games, it will be easier to obtain valuable insights which are helpful for enhancing user satisfaction when playing the game. In this study, we choose two games, Half-Life 2 with its two downloadable contents (HL2) and Half-Life: Alyx (HLA). We used the text analysis to discover topics concerned by players on online reviews of two video games. Different from traditional methods that may be affected by cognitive bias due to players' subjectivity (Yang et al. 2018), our approach enables us to get a comprehensive view of most players by analyzing the VR's impact on video games with a large-scale dataset.

RELATED WORK

Online review is one of the essential sources of product information for consumers, as well as a way for developers to understand user feedback on products. (Busurkina et al. 2020) used quantitative text analysis of game reviews on the Steam platform to understand the important characteristics of game experience. It supports the idea that

review analysis can help improve a product and increase sales efficiently. (Baowaly et al. 2019) evaluated the usefulness of game reviews on the Steam store. They collect user reviews from different game types and build a classification model to predict whether these reviews are useful.

In order to uncover the topics embedded in textual data, this study uses Latent Dirichlet Allocation (LDA) to discover the hidden topics that appear in reviews. LDA model is a generative topic model, which was first proposed by (Blei et al. 2003). It is a prevalent method of topic exploration and discussions. In applications of topic modeling for online reviews, previous works have shown that LDA is a significant tool to extract information embedded in review datasets. (Santos et al. 2019) used LDA to assess discrepancies in video game appraisal by experts and amateurs. (Yu et al. 2021) detected esports players' favorite topics by using the LDA method.

Immersion is a critical factor that VR games significantly advantage over desktop 3D games because VR techniques can provide a large field of view. (Seibert & Shafer 2018) investigated players using VR systems experienced significantly higher levels of spatial presence than using a non-VR system. (Yang et al. 2018) found most users reported a more intense experience for playing the game on the VR device and reported more "immersive" and fluency than playing games on a computer monitor.

Methodology

Latent Dirichlet Allocation (LDA) for Topic Modeling

This study applies Latent Dirichlet Allocation (LDA) (Blei et al. 2003) to discover the underlining topics in the review text. LDA assumes that each document in the corpus is represented as a random mixture of potential topics, and the feature of each topic is the distribution of words in a vocabulary. Due to the benefits that LDA can automatically determine the optimal numbers of topics and specify keywords or rules to infer each topic, we implement LDA for topic modeling in our Half-Life datasets. Figure 1 illustrates the graphical LDA procedure.

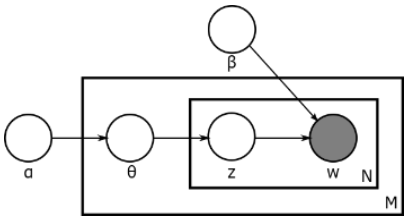

Figure 1. Graphical model representation of LDA. (Adapted from Blei et al. 2003)

Shapiro-Wilk test

Shapiro-Wilk test verifies whether a random sample, $\{x_1, x_2, ..., x_n\}$ comes from a normal distribution. It is calculated as follows:

$$W = \frac{(\sum_{i=1}^{n} a_i x_i)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

where x_i is the ordered sample values. a_i is the constant generated from the sample mean $\bar{x}$, variances and covariances of the order statistics of a sample of size n from a normal distribution.

Two-sample t-test

Two-sample *t*-test supposes that we would like to test the hypothesis that two samples have the same means, i.e., H_0: $\mu_0 = \mu_1$. In the following discussion, we assume that:

- both distributions are normal
- they have common variances

Thus, the two-sample *t*-test based on the hypotheses:

$$T = \frac{X_1 - X_2}{\sqrt{s_1^2/N_1 + s_2^2/N_2}}$$

where N_1 and N_2 are the sample sizes, X_1 and X_2 are the sample means, and s_1^2 and s_2^2 are the sample variances.

EXPERIMENTAL RESULTS

Dataset

We use the API provided by Steam to obtain a new dataset of the Half-Life series up to June 2021. The features of the data include a unique recommendation ID, user ID, review publication & updated date, the recommended / not recommended tag for the review, language tag, review text, the number of people who rated the review as helpful, and the number of people who rated the review as funny, and so forth. Before applying LDA topic modeling, basic text processing is implemented. The purpose of text processing is to eliminate non-English words, emoticons, or characters with insignificant meaning for information extraction. Stopwords removal is a process of removing words that are likely unimportant for information extraction such as “game”, “half”, “life”, “steam”, and “gg”. Then, the second step is tokenization, separating sentences into words. The last processing is stemming, reducing inflected words to their word base form. Table 1 presents the properties of our datasets after prepossessing.

Table 1: The properties of the datasets

Game	**HLA**	**HL2**
Positive	37,746	58,757
Negative	536	2,058

Total	38,282	60,815

Topic Modeling and Grouping with LDA

For finding the optimal number of topics, we analyze the coherence score of HLA, which is a measure used to evaluate a topic model (Newman et al. 2010). Figure 2 shows the number of topics from 3 to 100 with their coherence score, from that we could choose the optimal number of topics as 11 as the coherence score is flattened out after 11.

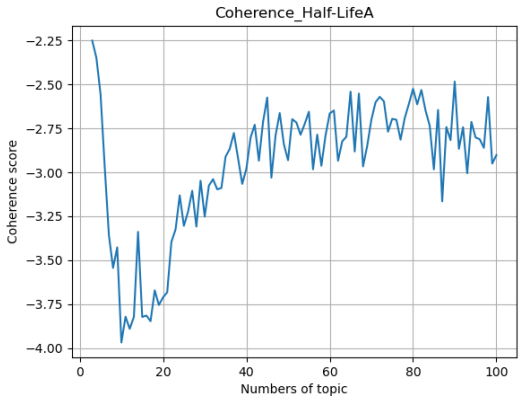

Figure 2. Coherence score of HLA.

We validate the inferred topics of the LDA model by reading the most weighted keywords of each topic and original review text to avoid implicit expressions or forum spam. Table 2 lists the topics we summarized from HLA reviews.

Table 2: Topics and related keywords

	inferred topic	**keywords example**
0	VR	“vr” *0.178 + “headset” *0.023…
1	gameplay	“gameplay” *0.008 + “mechanic” *0.005 …
2	graphics	“graphic” *0.009 + “detail” *0.003 …
3	optimization	“issue” *0.004 + “bug” *0.002 …
4	control	“movement” *0.005 + “controller” *0.004 …
5	plot	“story” *0.016 + “world” *0.007
6	character	“the Combine” *0.006 + “headcrab” *0.003 …
7	atmosphere	“environment” *0.005 + “horror” *0.005 …
8	combat	“weapon” *0.009 + “ammo” *0.004 …
9	content	“quest” *0.008 + “puzzle” *0.008 …

10	price	“money” *0.002 + “dollar” *0.001 …

Topic Prevalence and Statistics Comparison

In this stage, we feed those 11 topics inferred from the LDA model into the data analysis module to infer the topics’ popularity and positive rate among topics. We also compare topic positive between two games via hypothesis testing using a two-sample *t*-test to explore the impact of VR technology on video games.

In order to explore HLA players’ topic frequencies $\left(\frac{\#\text{of reviews contain the topic}}{\#\text{of total reviews}}\right)$, and topic positive rates $\left(\frac{\#\text{ of recommended reviews contain the topic}}{\#\text{ of total reviews}}\right)$ (Lin et al. 2019) for each specific topic, we summarize the frequency and positive rate of our inferred topics and present the result in Table 3.

Table 3: Prevalence and Sentiment Distribution of Topics

		Topics (%)		**Positive (%)**	
	Topics	HLA	HL2	HLA	HL2
0	VR	92.21%	0.50%	98.20%	92.74%
1	gameplay	12.02%	7.71%	96.48%	94.18%
2	graphics	12.70%	8.90%	96.60%	94.99%
3	optimization	5.57%	1.68%	90.30%	84.43%
4	control	8.95%	1.31%	94.72%	90.86%
5	plot	15.21%	16.76%	97.44%	94.83%
6	character	8.69%	7.14%	95.52%	91.94%
7	atmosphere	4.80%	1.76%	97.09%	95.97%
8	combat	12.51%	6.51%	93.04%	89.72%
9	content	5.11%	3.18%	94.23%	92.50%
10	price	3.03%	1.85%	94.08%	90.94%

The result of Table 3 shows VR-related reviews account for more than 90% of all reviews, and HLA players are at least 90% satisfied with each topic. Compared with other topics, the discussion on “VR” has an overwhelming advantage with a positive rate of 98.2%. Besides, the top five frequent topics of both games: “gameplay”, “graphics”, “plot”, “character”, and “combat” occupy approximately over half of all aspect terms mentioned in the reviews. Moreover, due to the new interactive method, the topic frequency of “gameplay”, “graphics”, and “combat” on HLA have increased by 3.8% ~ 6.0% compared with HL2. On the contrary, the topic frequency of “plot” does not change significantly. This result also illustrates the attractiveness of VR to players.

In addition, we choose these top five frequent topics and compare their positive rate. We calculate the *p-value* of the *Shapiro-Wilk* test to verify if they follow the

normal distribution. The result indicates both follow the normal distribution as shown in Figure 3.

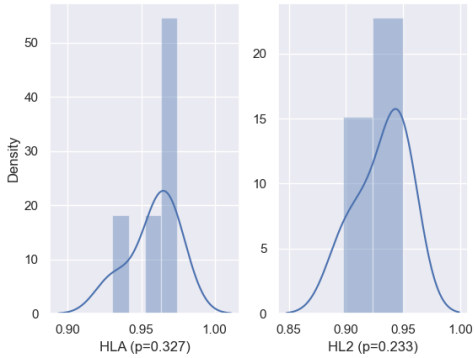

Figure 3. Positive rate distributions of top five topics.

Therefore, we perform a two-sample *t*-test with the significance level of 0.05 to see whether the positive rate means of the two games are equal or not. Because the *p-value* of our test (0.069165) is greater than $\alpha = 0.05$, we reject the hypothesis which assumes the two positive rate means are not equal. In other words, we do not have sufficient evidence to conclude that the positive mean rate between the two games is different.

CONCLUSIONS

This study aims to explore the impact of VR on players by analyzing reviews uploaded by players on Steam with data mining techniques. The main contribution of this paper is threefold. First, we used the topics modeling on about 100,000 Half-life series reviews from the Steam community. We uncover 11 topics embedded in HLA re-views which reflect essential VR game elements such as "VR", "gameplay", "graphics", "plot", and so forth. Second, we investigated the topic prevalence and positive rate among HLA reviews. Apart from "VR", we identified five topics that were most widely discussed by HLA players, which also attracted the most attention of HL2 players. Finally, we compared the HLA player positive rate with HL2 in five aspects extracted from the previous step. The result shows that although most HLA players are satisfied with VR experiences, VR has no significant impact on these game elements such as "graphics" or "plot". However, we believe that VR may positively influence the development of video games in the future.

There are apparent limitations of the current work. First, this study only considers English reviews, ignoring the enormous reviews from non-English reviewers. Moreover, the genre of VR games was limited. Other representative VR games such as *BONEWORKS* or *Beat Saber* were not analyzed, which might affect our recognition of emerging opinions where esports players are concern. Additionally,

game reviews are full of abbreviations and memes, which brings difficulties to our analysis. Our conclusions may suffer from these bias due to limited background knowledge.

REFERENCES

Alves, S., Callado, A. and Juc´a, P. (2020), "Evaluation of graphical user interfaces guidelines for virtual reality games", in *2020 19th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, IEEE, 71--79.

Baowaly, M. K., Tu, Y.-P. and Chen, K.-T. (2019), "Predicting the helpfulness of game reviews: A case study on the steam store", *Journal of Intelligent & Fuzzy Systems* 36(5), 4731--4742.

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003), "Latent dirichlet allocation", *the Journal of machine Learning research* 3, 993--1022.

Busurkina, I., Karpenko, V., Tulubenskaya, E. and Bulygin, D. (2020), "Game experience evaluation. a study of game reviews on the steam platform", In *International Conference on Digital Transformation and Global Society*, Springer, 117–127.

Lin, D., Bezemer, C.-P., Zou, Y. and Hassan, A. E. (2019), "An empirical study of game reviews on the steam platform", *Empirical Software Engineering* 24(1), 170--207.

Mimno, D., Wallach, H., Talley, E., Leenders, M. and McCallum, A. (2011), "Optimizing semantic coherence in topic models", In *Proceedings of the 2011 conference on empirical methods in natural language processing*, 262--272.

Newman, D., Lau, J. H., Grieser, K. and Baldwin, T. (2010), "Automatic evaluation of topic coherence", In *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*, 100--108.

Porter III, J., Boyer, M. and Robb, A. (2018), "Guidelines on successfully porting non-immersive games to virtual reality: a case study in minecraft", In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, 405--415

Santos, T., Lemmerich, F., Strohmaier, M. and Helic, D. (2019), "What's in a review: Discrepancies between expert and amateur reviews of video games on metacritic", *Proceedings of the ACM on human-computer interaction* 3(CSCW), 1--22.

Seibert, J. and Shafer, D. M. (2018), "Control mapping in virtual reality: effects on spatial presence and controller naturalness", *Virtual Reality* 22(1), 79--88.

Shelstad, W. J., Smith, D. C. and Chaparro, B. S. (2017), "Gaming on the rift: How virtual reality affects game user satisfaction", In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 61, SAGE Publications Sage CA: Los Angeles, CA, 2072--2076.

Yang, W., Rifqi, M., Marsala, C. and Pinna, A. (2018), "Towards better understanding of player's game experience", In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, 442--449.

Yu, Y., Nguyen, B.-H., Yu, F. and Huynh, V.-N. (2021), "Discovering topics of interest on steam community using an lda approach", In *International Conference on Applied Human Factors and Ergonomics*, Springer, 510--517.