Topic Modelling Applied to Negative Reviews of Elden Ring

Eduardo N. S. Ramos[1]

[1]Departamento de Engenharia Elétrica
Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ)

eduardonsantiago@aluno.puc-rio.br

Abstract. *User reviews in video games have been integral to digital distribution platforms since their inception, serving as crucial indicators for consumer decision-making by guiding potential players toward or away from specific titles. However, systematic analysis of gaming review topics and their distribution remains understudied in practical applications. This paper presents an approach to Topic Modeling applied to negative reviews of Elden Ring, utilizing the LLaMA language model for embedding generation, followed by dimensionality reduction through UMAP and clustering via HDBSCAN for topic identification. Our results demonstrate that contemporary Large Language Model architectures can generate semantically meaningful embeddings that effectively cluster similar issues, enabling the extraction of actionable insights from negative reviews.*

1. Introduction and Motivation

The video game industry has generated over 400 billion dollars in revenue as of 2024, surpassing the movie, book, and music industries [Statista 2024]. This form of entertainment differs significantly from others because it relies on user input rather than passive consumption. This means that each user's experience with a game is uniquely shaped by their interactions with the content. This interactivity allows for a personalized and dynamic engagement that is not typically found in other forms of media, where the experience is more uniform and passive.

This individual experience fosters video game communities, such as those on [Reddit 2024], where players can share their achievements, ask questions, and discover new ways to interact with games they are already familiar with.

In the case of *Elden Ring*, the game is known for its difficulty and overwhelming challenges compared to other games. Additionally, games released by the same company (*From Software*) possess atmospheric and dark fantasy elements that further reinforce a hostile game world. As pointed out by [Lachowski 2024], the world of *Elden Ring* is filled with almost insurmountable adversities and dark, decaying ruins, while still maintaining an element of hope, with non-player characters commenting on the player's chance to defeat the antagonists and achieve victory.

[Lachowski 2024] references this contrast with the element of 'Grace' in *Elden Ring*.

> The misfortune of being a Tarnished places us in the position of observers of the conflicts and their developments across the Lands Between, while

the contact with Grace conditions new transformations brought about by the battles with demigods and other secrets encountered along the way.

Regarding a user's main form of product evaluation in the video game industry, the Steam review system [FromSoftware 2024] allows players to create binary reviews (either positive or negative) of their played games, along with a written text detailing their impressions. Despite winning the Game Awards for Best Game of 2022 [IGN 2022], *Elden Ring* currently has around 92% positive reviews across all languages and reviews on Steam.

A system that can correctly classify written reviews based on semantic similarity Topic Modeling could be used to evaluate common themes among *Elden Ring* players, potentially highlighting user concerns and tracking complaint frequencies over time for specific topics. For example, this could reveal whether technical problem complaints decreased following patch releases.

2. Related Works

While research on topic modeling for game reviews is limited, several studies have employed techniques like BERTopic, which leverages transformer architectures to perform document clustering based on semantic similarity. Additionally, the literature includes investigations into sentiment analysis applications for gaming reviews, providing complementary approaches to understanding player feedback.

[Viggiato et al. 2022] and [Wang and Goh 2020] attempted to classify gaming reviews using sentiment analysis. The former concludes that analyzing text alone is problematic due to the prevalence of sarcastic reviews and bullet-point formats lacking contextual information necessary for classification models. In contrast, the latter achieves better results by incorporating topic modeling and extracting features such as word count and keywords directly from the text. However, neither study incorporates external information such as gaming hours to predict user reviews.

In [Silveira et al. 2021], the authors employed Legal-BERT to classify American legal documents and sentences. This approach successfully identified well-defined topics for characteristic sentences, such as those related to capital punishment, women's rights, and copyright. However, it struggled to classify other types of crimes and sentences effectively.

In [Egger and Yu 2022], the authors conducted a comparative analysis of traditional topic modeling techniques, such as Latent Dirichlet Allocation (LDA), with more modern approaches like Top2Vec and BERTopic. These methods were applied to Twitter posts related to the COVID-19 pandemic. The analysis demonstrated that classical topic modeling techniques performed poorly in domains with a high degree of response diversity. In contrast, more robust methods like BERTopic and Top2Vec successfully identified topics but produced an excessively large number of groups, rendering the final results nearly uninterpretable.

3. Methodology

This section will explore the methodology employed for dataset preparation, the generation of text embeddings, dimensionality reduction, and clustering. It will also detail the

metrics used for hyperparameter tuning and the approach taken for the final evaluation of the identified topics.

3.1. Dataset

The dataset was obtained using Steam's Web API [Valve 2024], which provided users' Steam IDs and review information, as summarized in Table 1. This table includes the review's recommendation variable, 'voted_up', along with the total number of reviews and the number of games owned by each user.

steamid	playtime_forever	num_reviews	num_games_owned	voted_up
76561199098034787	303	13	0	True
76561199013405597	82	12	0	False
76561199173536454	437	1	0	True
76561199118169519	88	2	0	False
76561198809081703	4957	2	18	True

Table 1. Reviews Dataset

3.2. Topic Modelling

Topic modeling is an unsupervised machine learning technique designed to automatically uncover latent themes or topics within large collections of documents. This is achieved by analyzing patterns of word co-occurrence [Blei et al. 2003] and capturing semantic relationships between terms [Angelov 2020]. By identifying groups of words that frequently appear together, topic modeling provides a representation of the underlying themes in textual data, enabling a structured analysis from an unstructured dataset.

3.3. Text Embedding

The text embeddings were generated using the Llama 3.2-1B model [Dubey et al. 2024], which was run locally on an RTX 3060 GPU with 12 GB of VRAM, 32 GB of RAM, and an AMD Ryzen 5 5600X CPU. The process involved passing each negative review through the transformer model and extracting the values from the final hidden layer, which has a dimensionality of 512. These embeddings capture the semantic similarity of each review's text and serve as the input for subsequent clustering steps.

3.4. Clustering

The topic modeling process involved a clustering step applied to the embeddings generated from negative reviews. Each embedding was treated as a vector in a high-dimensional space, with similar reviews grouped based on their semantic similarity. Before clustering, a dimensionality reduction step was performed using Uniform Manifold Approximation and Projection (UMAP) [McInnes et al. 2020]. UMAP was selected for its ability to preserve the original topological structure of the data, which aids in forming semantically coherent groups.

The clustering itself was conducted using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) [Malzer and Baum 2020]. This

method was chosen due to its ability to generate hierarchical clusters that can be fine-tuned to suit the specific requirements of the problem. HDBSCAN identifies core, frontier, and outlier points within the dataset, making it robust enough to discard reviews that do not fit into any meaningful group. This characteristic ensures that only relevant and well-defined clusters are retained, enhancing the interpretability of the topic modeling results.

3.4.1. Hyperparameter Tuning

3.4.2. Hyperparameters

To identify an optimal combination of hyperparameters for the dimensionality reduction and clustering models, an optimization strategy was employed. This approach involved systematically varying the hyperparameters listed in Tables 2 and 3, corresponding to UMAP and HDBSCAN, respectively. The goal was to evaluate different parameter configurations to determine the most effective settings for achieving meaningful and interpretable clustering results.

Hyperparameter	Values
n_neighbors	5, 25 and 50
n_components	2, 5 and 8
min_dist	0, 0.1, 0.3 and 0.5
metric	euclidian and cosine

Table 2. UMAP Hyperparameters Used in Optimization

Hyperparameter	Values
min_cluster_size	5, 25 and 50
min_samples	1, 5 and 10
cluster_selection_epsilon	0, 0.5 and 1
cluster_selection_method	eom and leaf

Table 3. HDBScan Hyperparameters Used in Optimization

The UMAP optimization parameters included adjustments to the number of neighbors considered for each point during dimensionality reduction, the target dimensionality, the minimum distance between points in the reduced space, and the choice of distance metric. For HDBSCAN, the hyperparameters involved modifications to the minimum cluster size and the balance between local and global clustering. These adjustments aimed to identify the parameter configurations that best represented topics in a manner consistent context of this paper.

3.4.3. Metrics

The metrics employed for optimization included the Silhouette Score [Rousseeuw 1987], the Calinski-Harabasz Index [Caliński and Harabasz 1974], and the Davies-Bouldin Index [Davies and Bouldin 1979]. The Silhouette Score evaluates the distinctiveness of

each cluster by measuring the similarity of points within a cluster compared to points in other clusters, effectively assessing how well-separated the clusters are. The Calinski-Harabasz Index and Davies-Bouldin Index measure the ratio of intra-cluster dispersion to inter-cluster separation, capturing the structural integrity of clusters relative to each other.

Since these metrics operate on different scales, their values were normalized to a range between 0 and 1. The optimization process aimed to maximize the mean value across the three metrics.

4. Results

The topic modeling analysis was conducted on a randomly selected sample of 5,000 reviews, resulting in three distinct topic groups, along with an additional outlier group. The distribution and characteristics of these groups are summarized in Table 4. The clustering achieved a Silhouette Score of 0.38, indicating moderate separation between the identified clusters, suggesting that while the clusters exhibit some degree of uniqueness, there is room for further refinement in distinguishing them.

Topic	N of Reviews
Technical Problems	1451
Constructive Criticisms	1367
Non-constructive Criticisms	906
Outliers	1276

Table 4. Topic Modelling Results

Representative samples of 5 reviews per topic are presented in Tables 5, 6, and 7. These tables illustrate distinct categories of criticism: non-constructive feedback, which offers minimal actionable insights for development teams; constructive criticism, which provides comprehensive assessments identifying specific areas for improvement; and technical issues, which highlight performance and stability concerns. The constructive criticism cluster synthesizes key factors contributing to negative reviews, thereby identifying potential areas for enhancement. The technical issues cluster specifically addresses system performance and gameplay stability concerns.

Non-constructive Criticisms
”it’s an impossible game. You cannot play because you will meet a big boss and cannot pass.”
”would rather play modded DS3 than this.”
”dog ass compared to DS2.”
”got called maidenless and quit. This game is shit.”
”the sup is bad.”

Table 5. Table of Non-constructive Criticisms

Constructive Criticisms
“Most of the game past the first two areas, and especially past the Capitol, is rehashed, copy-pasted content with almost nothing unique”
“The open world with tons of copy and paste dungeons, enemies, and bosses, you just fight different colors of them over and over.”
“I’ve been playing this game offline MODDED the way it should of been released. The original game has terrible a gatekeeping crafting system that forces people to use CHEAT ENGINE and exploits to dupe items.”
“the story is supposed to be vague and implied but coming off of the atmosphere of DS this just feels like the same reheated mess but less interesting”
“Endgame feels like a large leap from Late-Game in terms of difficulty, and you are often better not engaging in field bosses.”

Table 6. Table of Constructive Criticisms

Technical Problems
“have experienced 2 crashes in the first hour of gameplay and a number of massive FPS drops into the single digits”
“played for 6 hours yesterday and logged in today and brought a mismatched save files between steam cloud and my local machine and bam progress is ALL gone.”
“Controller does not work!”
“It crashes when trying to do multiplayer a lot.”
“Poorly optimized for PC, framerate spikes and freezes for no reason. crashes during basic encounters.”

Table 7. Table of Technical Problems

Subsequently, the same model and optimization process was applied exclusively to the Constructive Criticism cluster, yielding a silhouette coefficient of 0.3 and revealing two distinct groups of concerns. The first group centered on the game’s overall scope and completion time, while the second focused on difficulty scaling. The results are presented in Table 8, while a sample of 5 reviews is shown in Tables 9 and 10.

Topic	N of Reviews
Difficulty	559
Game Scope	236
Outliers	572

Table 8. Topic Modelling Results

Game Scope
"The way bosses are lined up and how that line can be broken thanks to the open world setting"
"the big open world doesn't really add anything to the experience. It just means you have to monotonously ride around on your horse from one level to another."
"The open world makes this game take 10x longer than it should"
"DO NOT purchase this game. If you cannot dedicate hours a day to mastering the wildly brutal combat this game provides"
"Elden Ring is boring and repetitive to death, with a generic open world that gets old very fast."

Table 9. Table of Reviews Related to Game Scope

Difficulty
"I get one hit death because the boss is only beatable when you roll around it in your trousers."
"This game is very difficult and has a huge learning curve I would not recommend"
"Plain and simple. This is a very hard game. You will die, die, and die forever more. If you're looking for a fun game, this is definitely not it."
"I felt I constantly needed to go on the internet to find something, figure how to do something, figure a quest out and so on"
"If you've got a lot of time on your hands then possibly a worthy buy on discount but other than that avoid this game."

Table 10. Table of Reviews Related to Game Difficulty

5. Conclusion and Future Works

The implementation of Topic Modeling utilizing LLaMA coupled with clustering algorithms yielded effective results in identifying common thematic patterns among *Elden Ring* players' feedback. The initial clustering analysis revealed that a substantial portion of players encountered legitimate concerns regarding the game's difficulty and scope, alongside technical issues. Given that this study focused on the PC player base, which experiences greater hardware heterogeneity compared to console platforms, the prevalence of technical issues suggests potential optimization inadequacies during development.

Further analysis of constructive criticism revealed that game difficulty was a predominant concern, while criticism regarding game size constituted a smaller but distinct cluster with numerous statistical outliers. These findings suggest that while critical feedback is diverse, and the game's difficulty aligns with the intended design philosophy, the expansive open-world structure presents a notable challenge. Players reported spending disproportionate time traversing the environment rather than engaging with core gameplay mechanics.

Future research directions could explore the application of alternative Large Language Model architectures for Topic Modeling, particularly in decomposing primary topics into more granular subtopics of common issues. Additionally, investigating the effec-

tiveness of linked prompting strategies, where topic modeling generates common issues and subsequently categorizes individual reviews, presents another promising avenue for development.

References

Angelov, D. (2020). Top2vec: Distributed representations of topics.

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

Caliński, T. and Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27.

Davies, D. L. and Bouldin, D. W. (1979). A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2):224–227.

Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., Mathur, A., Schelten, A., Yang, A., Fan, A., et al. (2024). The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Egger, R. and Yu, J. (2022). A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. *Frontiers in sociology*, 7:886498.

FromSoftware (2024). Elden ring steam page. `https://store.steampowered.com/app/1245620/ELDEN_RING/` [Accessed: 2024-05].

IGN (2022). The game awards 2022 winners: The full list. `https://www.ign.com/articles/the-game-awards-2022-winners` [Accessed: 2024-05].

Lachowski, V. F. (2024). Desgraça em elden ring: Expressões alegóricas de ruínas e cadáveres.

Malzer, C. and Baum, M. (2020). A hybrid approach to hierarchical density-based cluster selection. In *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, page 223–228. IEEE.

McInnes, L., Healy, J., and Melville, J. (2020). Umap: Uniform manifold approximation and projection for dimension reduction.

Reddit (2024). r/ elden ring. `https://www.reddit.com/r/Eldenring/` [Accessed: 2024-05].

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

Silveira, R., Fernandes, C. G., Araujo Monteiro Neto, J., Furtado, V., and Pimentel Filho, J. E. (2021). Topic modelling of legal documents via legal-bert.

Statista (2024). Video game industry - statistics facts. `https://www.statista.com/topics/868/video-games/#topicOverview` [Accessed: 2024-05].

Valve (2024). Steam web api documentation. `https://steamcommunity.com/dev` [Accessed: 2024-05].

Viggiato, M. et al. (2022). What causes wrong sentiment classifications of game reviews? *IEEE Transactions on Games*.

Wang, X. and Goh, D. H.-L. (2020). Components of game experience: An automatic text analysis of online reviews. *Entertainment Computing*, 33:100338.