

# LLMを用いたゲームレビューの分析

坪井広樹(株式会社NTTデータグループ) 古市佑哉(富士通株式会社) 渡邊宏太郎(株式会社クレスコ)

## ゲームレビューの分析手法における問題点

従来の分析手法ではレビュー評価の傾向を踏まえた分析が難しい。

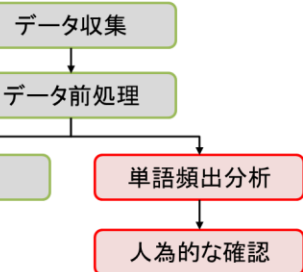
- ①目視によるレビューテキストの確認を行うため、大量のレビューテキストデータを処理できない。
- ②単語の頻度のみを分析しているため、文脈に依存するニュアンスの理解や同義語の考慮が難しい。

## LLMを用いた分析手法の提案による解決

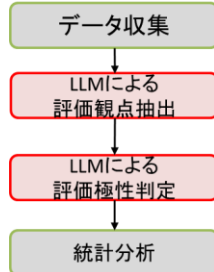
- ①大量のデータを短時間で解析できることで、時間を97.1%削減できた。
- ②単語頻度による評価の限界を克服し、皮肉や婉曲表現などのニュアンスを適切に捉えて解析を行うことができた。

## 既存手法・提案手法

### 既存手法



### 提案手法



既存手法の単語頻出分析、人為的な確認をLLMで代替。  
レビューテキストを評価するための評価観点をLLMを利用して抽出。

## 評価方法

- ①人手による評価極性判定と提案手法の比較  
⇒GPT-4oによる評価極性の判定結果と人手による評価極性の判定結果を比較し、LLMの精度を確認する。  
また、分析時間の比較も行う。
- ②評価極性の分布比較  
⇒提案手法で収集したレビューテキストの評価データからマン・ホイットニーU検定を用いることで日本語のレビューと英語のレビューについてポジティブ・ネガティブ評価の分布比較を行う。

## 評価結果

### ①人手による評価極性判定と提案手法の比較

評価観点	一致率(%)
gameplay	75.0
bugs	89.6
difficulty	85.4
controls	93.8
performance	93.8
story	89.6
remake	97.9
online	85.4
accessibility	91.7
characters	89.6
post_game	100.0
worldview	95.8
Price	89.6
replayability	77.1
community	91.7
modes	91.7
added features	93.8
volume	97.9

評価観点	一致率(%)
immersion	85.4
freedom	91.7
sound	100.0
graphic	89.6

	1件あたりの 分析時間(分)
人手	4.12
LLM	0.12

全22件の評価観点のうち20件で一致率が85%以上となったが、残りの2件は他と比較して10%程度一致率が低くなった。  
分析時間の比較では1件あたりの分析時間は97.1%削減した。

### ②評価極性の分布比較

	U_statistic	p_value	Significant
bugs	7076686	3.64E-30	Yes
difficulty	7090111	5.30E-25	Yes
controls	6823623	2.00E-14	Yes
performance	6793134	1.21E-09	Yes
worldview	6363703	4.54E-04	Yes

日本語レビューは比較的ネガティブに評価する傾向があった。特にバグ、適正な難易度、操作性などの評価項目についてその傾向が強い。ゲームの世界観を評価する項目のみ英語レビューの方がネガティブに評価する傾向となった。

## 考察・課題

### ■評価傾向

- ・日本人がより厳しい評価をする傾向があると考えられる。
- ・日本人はプレイ体験の快適さを求める一方で海外の方はゲームの世界観を求める傾向があることも示唆される。

### ■今後の課題

- ・多言語への応用
- ・LLMの分析精度の向上