

RESEARCH ARTICLE

Mining Insights From Esports Game Reviews With an Aspect-Based Sentiment Analysis Framework

YANG YU¹, DUY-TAI DINH^{2,3,4}, BA-HUNG NGUYEN⁵, FANGYU YU¹,
AND VAN-NAM HUYNH¹, (Member, IEEE)

¹School of Knowledge Science, Japan Advanced Institute of Science and Technology (JAIST), Nomi 923-1211, Japan

²Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

³International School, Duy Tan University, Da Nang 550000, Vietnam

⁴The Kyoto College of Graduate Studies for Informatics (KCGI), Kyoto 606-8225, Japan

⁵Faculty of Economics and Management, Thai Binh Duong University, Nha Trang, Khanh Hoa 650000, Vietnam

Corresponding author: Yang Yu (yangyu@jaist.ac.jp)

ABSTRACT The explosive growth of player-versus-player games and tournaments has catapulted esports games into a rapidly expanding force in the gaming industry. However, novice and armature players' voices are often inadvertently overlooked because of a lack of effective analytical methods, despite the close collaboration between professional esports teams and operators. To ensure the quality of esports game services and establish a balanced gaming environment, it is essential to consider the opinions of unprofessional players and comprehensively analyze their reviews. This study proposes a new framework for analyzing esports reviews of players. It incorporates two key components: topic modeling and sentiment analysis. Utilizing the Latent Dirichlet Allocation (LDA) algorithm, the framework effectively identifies diverse topics within reviews. These identified topics were subsequently employed in a prevalence analysis to uncover the associations between players' concerns and various esports games. Moreover, it leverages cutting-edge Bidirectional Encoder Representations from Transformers (BERT) in conjunction with a Transformer (TFM) downstream layer, enabling accurate detection of players' sentiments toward different topics. We experimented using a dataset containing 1.6 million English reviews collected up to December 2021 for four esports games on Steam: TEKKEN7, Dota2, PUBG, and CS:GO. The experimental results demonstrated that the proposed framework can efficiently identify players' concerns and reveal interesting keywords underlying their reviews. Consequently, it provides precise insights and valuable customer feedback to esports game operators, enabling them to enhance their services and provide an improved gaming experience for all players.

INDEX TERMS Esports, topic modeling, prevalence analysis, sentiment analysis, steam.

I. INTRODUCTION

Esports is a rapidly growing industry that supports over two billion players and spectators worldwide, creating a billion-dollar revenue market annually. The competitive characteristics of esports have attracted interest from players, as these games emphasize Player versus Player (PvP) rather than just graphics and storylines. This makes it more enjoyable to watch competition. Recently, esports were debuted as an official medal sport in the 2022 Asian Games [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Chuan Li.

Esports games can be classified into several groups: first-person shooter (FPS), battle royales (BR), multiplayer online battle arena (MOBA), real-time strategy (RTS), fighting, and card games. According to the global esports market investigation by Newzoo [2], for the top 25 live esports games watched on YouTube, Twitch, and Mixer, 72% of the games consist of FPS, BR, MOBA, and Fighting. These games are typically much more hardcore because of their competitive elements. By analyzing over 100 million users on Steam, Baumann et al. [3] demonstrated that hardcore gamers can be grouped into separate clusters, such as FPS [4], action game [5], and Dota 2 [6], which match our observation with features of esports games.

As new games are released for desktop gaming clients, such as Steam, Origin, Uplay, and Battle.net, these clients have developed rapidly in recent years. They allow players to purchase and enjoy one-stop gaming services. Future updates can be easily made, eliminating players needing to find download paths and update the games themselves. With the development of clients, gaming communities have grown gradually, generating forums such as Steam, Ubisoft, Origin, and IGN. Steam is the ultimate destination for playing, discussing, and creating games among these platforms, containing nearly 30,000 games, from AAA publishers to mid-sized or major publishers. Steam has a large community with over 100 million potential players and provides a comprehensive and large-scale review system. Several previous studies [7], [8] collected data from Steam to analyze player behavior.

In the digital transformation era, data can easily be collected from various sources in diverse formats, with text being one of the most common unstructured data types. However, traditional technologies have become unable to handle such massive and unstructured data, leading to the emergence of big data in text classification, cutting-edge machine learning, and deep learning techniques that can perform highly accurate lower-level engineering and computation functions. Text classification is at the heart of various software systems that process textual data on a large scale. It performs fundamental natural language processing (NLP) tasks and is broadly applied for sentiment analysis, spam detection, and topic labeling.

Sentiment analysis [9] is an important topic in NLP. It aims to systematically and automatically analyze the opinions and emotions contained in text. Sentiment analysis classification algorithms can detect positive or negative sentiments in a text at different levels, such as document, sentence, and aspect levels. Users may have different granularities in interesting aspects [10]. Thus, aspect-based sentiment analysis (ABSA) is more sophisticated in reflecting users' detailed opinions and sentiments. An ABSA task involves two essential sub-tasks: aspect term extraction (ATE), which involves identifying the aspects of interest, and aspect sentiment classification (ASC), which involves determining the polarity of identified aspects [11], [12], [13], [14].

Lin et al. [7] conducted a study on Steam reviews and found that game reviews have unique characteristics compared with other types of reviews. However, the analysis of game reviews presents several challenges. First, Steam game reviews use the "Recommended" or "Not Recommended" system instead of a rating system on a scale of 1 to 5. This system may not always accurately reflect the author's attitude because of the possibility of incorrect selection. Second, game reviews are often unhelpful, with Lin et al. [7] identifying that 58% of reviews fall into the "Not Helpful" category. This makes it difficult to find informative reviews without the help of analytical tools. Valuable game reviews can typically be summarized into categories, such as Pros, Cons, Bugs, Suggestions, and Videos [7]. Thus, using ABSA can be useful

in reducing the workload and enabling esports operators to grasp key information quickly and accurately.

Recently, the attention of many researchers has been drawn to sentiment analysis of game clients [15], [16], [17]. Game reviews on game clients serve as valuable resources to reflect esports enthusiasts' preferences and existing problems, allowing operators to enhance the quality of an esports game [18]. For instance, Hyeong et al. [15] showed that the imbalances between players of different ranks resulting from the services offered by esports operators can cause a massive loss of noob players. Conversely, such imbalances can make the game too easy, which contradicts the nature of esports games. Professional and noob esports players are the two ends of a scale and both are critical to the future of an esports game. However, esports operators prioritize professional players' opinions when serving these two groups, often ignoring large but silent noob players [16]. These observations motivated the design of an approach to help esports game operators to gain more accurate information and feedback from the enormous number of middle/low-ranked players to enhance their services. The major contributions and innovations of this study are as follows:

- We collected a large-scale dataset from the Steam platform, including four representative esports games: TEKKEN 7 (Fighting), Dota2 (MOBA), PUBG (BR), and CS:GO (FPS). All English reviews for these games up to December 2021 on Steam were collected. We identified the topics mentioned in the esports reviews using the Latent Dirichlet Allocation (LDA) algorithm. These keywords were then used to annotate 3,000 sentences in the dataset using three experienced annotators. The annotated dataset was used to fine-tune the training model on the input dataset with different downstream layers to achieve the best sentiment analysis performance. Our hybrid model takes advantage of unsupervised and supervised learning to improve the quality of sentiment analysis.
- We conducted an extensive experiment on the collected datasets to evaluate the performance of the proposed framework in terms of sentiment analysis quality. We used various metrics, such as precision, recall, micro F1, and macro F1. Our experimental results show that the attitudes of esports players toward different topics can reveal the future direction of esports game operators.

The remainder of this paper is organized as follows. Section II reviews previous studies on aspect-based sentiment analysis, topic modeling, and research on game reviews. Section III introduces the proposed framework. Section IV presents experimental results and analysis. Section V concludes the paper and outlines directions for future research.

II. RELATED WORK AND PRELIMINARIES

A. ASPECT-BASED SENTIMENT ANALYSIS

Aspect-based sentiment analysis (ABSA) is a technique used to measure users' sentiments toward specific aspects of

a product or service based on their comments on online forums. In the case of esports games, players consider multiple aspects such as graphics, storyline, and sociability. For instance, if the target aspect is graphics, a sentence like “the game graphics are very good” would have a positive sentiment. However, sentiment would be negative if the target aspect were cheaters.

Traditionally, ABSA tasks have been divided into aspect extraction and aspect sentiment classification. The former aims to identify the aspects of reviews that players focus on, and can be performed through supervised or unsupervised learning approaches. Pontiki et al. [19] provided multiple annotated datasets from seven domains and eight languages for an ABSA task. Yauris and Khodra [20] proposed a modified Double Propagation method to categorize aspect terms into groups, producing an aspect-based summary. Straat et al. [18], [21] analyzed consumer attitudes toward video games by extracting features based on word frequency and manually judging the polarity.

Recently, Baowaly et al. [22] designed a gradient-boosting algorithm to identify helpful and bad reviews from Steam game reviews. They also built a regression-based model to predict the review scores. Other studies have used deep learning techniques [23], [24], but all require annotated datasets to train classifiers.

Although single-method approaches have been used to analyze game reviews, they may not capture all the different aspects of the reviews, which can affect the accuracy of classification. To address this limitation, researchers have developed a more comprehensive approach called End-to-End Aspect-based Sentiment Analysis (E2E-ABSA). This unified model combines aspect extraction and sentiment classification, resulting in a holistic and accurate analysis of game reviews. The goal of E2E-ABSA is to detect the aspects and their corresponding sentiments. Ma and Hovy [25] proposed an E2E-ABSA approach that combined LSTM, CNN, and CRF. Schmitt et al. [26] compared different neural network models such as LSTM and CNN, for ABSA tasks. Li et al. [27] developed and evaluated varying downstream layers as a replacement for BERT’s original output layer for the E2E-ABSA task. Agarwal and Sabharwal [28] applied an end-to-end pipeline to classify tweets. Compared to separate ABSA subtasks, E2E-ABSA requires no feature engineering or data preprocessing, making it a promising approach for practical use.

B. TOPIC MODELING FOR ONLINE REVIEWS

The Latent Dirichlet Allocation (LDA) algorithm, introduced by Blei et al. [29] has been instrumental in discovering latent topics in large corpora of text data. For instance, Huang et al. [30] used LDA to identify various themes in Yelp reviews of restaurants, including service, value, decoration, and health. Feuerriegel et al. [31] analyzed the impact of topics in corporate press releases on stock market returns, whereas Santos et al. [32] applied LDA to assess disparities in video game evaluations by experts and amateurs.

Heng et al. [33] discovered that Amazon customer reviews of food products are influenced by four key factors: Amazon Service, Physical Features, Flavor Features, and Subjective Expression. Similarly, Putri et al. [34] used LDA to extract specific topics from travel reviews, while Tran et al. [35] analyzed TripAdvisor reviews to uncover 11 major topics, including food, hotel facilities, price, and staff.

C. GAME REVIEWS

Compared to platforms such as Twitter, which impose restrictions on the number of words and symbols, game platforms have little to no restrictions on game reviews, resulting in uneven quality of comment data and invalid expressions. Game reviews may contain invalid information such as special characters or spam messages. Furthermore, most game platforms do not restrict language, leading to a lack of standardization of the review text information obtained from game platforms. Consequently, research on game reviews differs from that on other platforms.

Over the past two decades, several studies have been conducted on game reviews for game user research. Gifford [36] analyzed the differences in reviews between video games and films. Lin et al. [7] pointed out that game reviews differ from mobile app reviews in several aspects, and both positive and negative reviews could be useful to game operators. Zagal et al. [17] demonstrated the relationship between game rating and sentiment words chosen by players. Bond and Beale [37] identified the features of a good game by analyzing game reviews. Livingston et al. [38] pointed out that game reviews and ratings affect commercial success.

Most research on game reviews focuses on game rating and selection and considers esports games as common as other games, but lacks an independent analysis of esports game review contents. Therefore, this study aims to explore esports players’ potential opinions using an E2E-ABSA model based on BERT [27], trained on a new annotated dataset for reviews from the Steam community. The obtained topics and sentiments support esports game operators and platforms in gaining deeper and useful information and feedback to enhance their services.

TABLE 1. Table of notations and abbreviations.

Symbol	Description
X	The input sentence
T	The length of the input sentence
k	The number of topics
A	The embedding input
E2E-ABSA	End-to-End Aspect-based Sentiment Analysis
BERT	Bidirectional Encoder Representations from Transformers
LDA	Latent Dirichlet Allocation
TFM	Transformer
GRT	Game-related Topics
PRT	Player-related Topics

III. PROPOSED FRAMEWORK

A. THE WORKFLOW

Proposed framework for topic prevalence analysis and aspect-based sentiment analysis of esports game reviews are shown

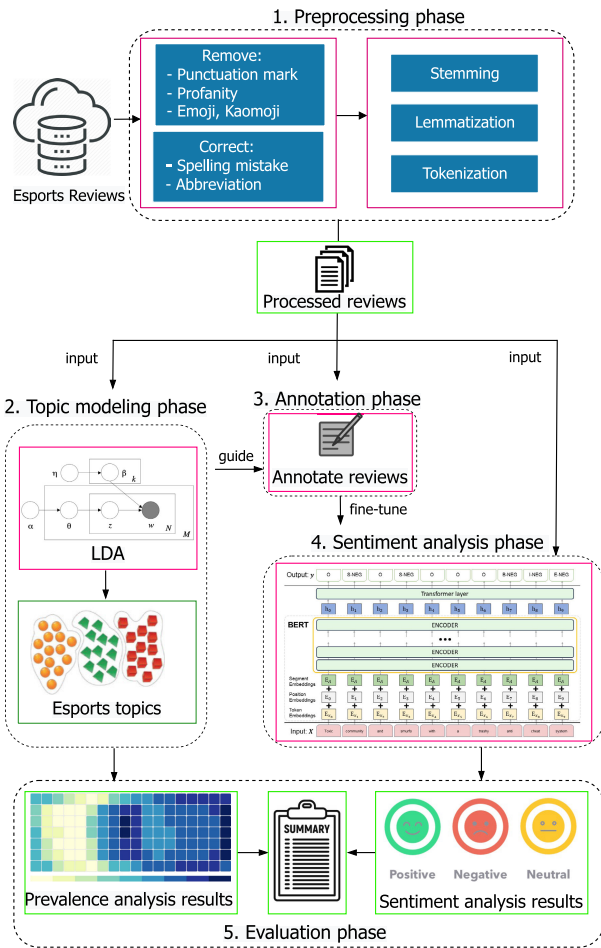


FIGURE 1. The workflow of the proposed framework.

in Fig. 1. The framework consisted of five phases. In the first phase, a new dataset of esports game reviews was collected from Steam and preprocessed using techniques such as key information detection and extraction (e.g., user’s SteamID, updated date, number of helpful and funny votes, language tag and reviews), language detection and filtering, noise removal, and spelling correction. In the second phase, Latent Dirichlet Allocation (LDA) was applied to perform topic modeling on the preprocessed dataset. This allowed for automatic discovery of topics and keywords for each esports game. In the subsequent phase, the identified topics were used for prevalence analysis and annotation guidance. In the third phase, three experienced annotators were invited to label aspects of each review based on the topics and keywords identified in the previous phase. The annotated dataset was then used to fine-tune the sentiment analysis model in the next phase. In the fourth phase, a BERT-based model was trained with a Transformer layer as the downstream layer to perform sentiment analysis tasks best. In the last phase, we analyzed their frequencies to infer each topic’s prevalence in the four esports games. In addition, we analyzed the sentiment polarity of each topic using our trained BERT model. Finally, we drew conclusions based on the obtained results for

the entire dataset. The proposed framework consists of data collection and preprocessing, topic modeling, aspect-based sentiment analysis, result visualization, and analysis.

B. LATENT DIRICHLET ALLOCATION (LDA)

In this study, we employed Latent Dirichlet Allocation (LDA) [29], a commonly used topic modeling technique, to uncover the underlying topics in the review text. LDA is a generative probabilistic model that 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. To determine the optimal number of topics, LDA iteratively performs topic and word distribution analysis, calculates coherence scores with different numbers of topics, and compares their values. In our study, we utilized LDA on the experimental dataset to identify the optimal number of topics and specify keywords and rules for inferring each topic. The pseudocode for the LDA algorithm is presented in Algorithm 1:

Algorithm 1 Generative Process of LDA [29]

- 1: **for** each topic $k \in [1, K]$ **do**
- 2: sample mixture components $\phi_k \sim Dir(\beta)$
- 3: **end for**
- 4: **for** each document $d \in [1, D]$ **do**
- 5: sample mixture proportion $\theta_d \sim Dir(\alpha)$
- 6: **for** each word $n \in [1, N_d]$ **do**
- 7: sample topic index $z_{d,n} \sim Mult(\theta_d)$
- 8: sample a term for word $w_{d,n} \sim Mult(\phi_{z_{d,n}})$
- 9: **end for**
- 10: **end for**

C. BERT-BASED MODEL ARCHITECTURE

1) BERT AS EMBEDDING LAYER

BERT is a powerful language representation model developed by Google [39] as illustrated in Fig. 2. BERT leverages masked language models to pre-train deep bidirectional representations, achieving impressive performance on sentence and token-level tasks. Given BERT’s exceptional feature capture capabilities, we used it as the embedding layer to extract sentence information. Specifically, we utilized the pre-trained “bert-base-uncased model”¹ with default parameter settings, such as num_hidden_layers (number of hidden layers in the Transformer encoder) set to 768 and the max_position_embeddings (the maximum sequence length that this model can be used with) set to 512. These parameters can be adjusted to suit the size of the model.

To ensure that the review text adhered to BERT’s input-size limitations, we checked the token size of each review and split any overlength reviews into several paragraphs. We also manually reviewed all feedback and removed any instances of ASCII art steam reviews to obtain a cleaner dataset for input. For a given sentence $X = (x_1, x_2, \dots, x_t)$, where

¹<https://github.com/huggingface/transformers>

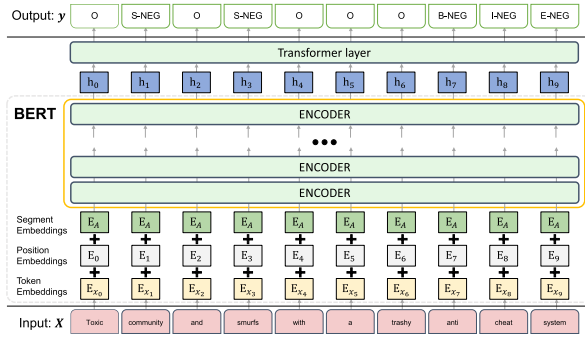


FIGURE 2. Overall of the E2E-ABSAs model architecture.

$1 \leq t \leq 512$ is the length of the sentence, the embedding space is represented by vectors encapsulating the meaning of each word, with similar words having closer vector values. BERT’s input embeddings are packaged as $A = (a_1, a_2, \dots, a_t)$, which is the sum of the token embeddings, the segmentation embeddings, and the position embeddings. The remainder of the implementation is virtually identical to that of the original BERT, and we refer to the transformer layer without delving into an in-depth description of the model architecture.

2) DESIGN OF DOWNSREAM LAYER

Once we obtained the BERT representations, we added a transformer layer as the final layer to fine-tune the BERT model with various designs for the E2E-ABSAs task. The Transformer model, proposed by Vaswani et al. [40] provides a robust feature extractor while discarding traditional RNN architectures in NLP tasks. Additionally, since BERT is fundamentally made up of the encoder of the Transformer model, we used a Transformer layer with the same architecture as the BERT encoder in this study. The TFM’s computational process is straightforward and can be broken down into several steps, as shown below:

$$Q_i, K_i, V_i = (W_i^Q, W_i^K, W_i^V)A \quad (1)$$

$$\text{Att}_i(Q, K, V) = \text{softmax} \left(\frac{Q_i K_i^T}{\sqrt{d_k}} \right) V_i \quad (2)$$

$$\text{Self-Att}_{MH}(Q, K, V) = \text{Concat}(\text{Att}_i) W^O \quad (3)$$

$$\hat{A} = \text{LN}(A + \text{Self-Att}_{MH}(Q, K, V)) \quad (4)$$

$$y = \text{LN}(\hat{A} + \text{FFN}(\hat{A})) \quad (5)$$

In the above equations, $Q, K,$ and V represent the Query, Key, and Value vectors that are generated from the input vector. The feed-forward network (FFN) consists of a simple fully-connected neural network that uses ReLU as the activation function [40]. Formulas 1-3 illustrate the computational process of the multi-head self-attention mechanism with i heads, while Formulas 4-5 show the residual connection step [41] and layer normalization. Finally, a linear layer with

softmax activation is attached to the output of the TFM layer to obtain the final prediction.

D. TOPIC PREVALENCE ANALYSIS

We employed a data analysis approach that combined topic words discovered from a topic model with a manual review of the text to determine the prevalence of each topic in esports games. To calculate the topic prevalence (TP) for each game, we used the following equation:

$$\text{TP} = \frac{\# \text{ of reviews which contain the specific topic}}{\# \text{ of total reviews}} \quad (6)$$

To ensure the accuracy of our analysis, we excluded keywords with multiple meanings from the frequency analysis process to avoid potential statistical bias. It is worth noting that since comments may cover multiple topics and each topic may have multiple interpretations, the frequency analysis results are more qualitative than quantitative.

TABLE 2. The properties of the processed datasets.

Game	TEKKEN7	Dota2	PUBG	CS:GO
#Recommended	24,906	397,090	189,776	726,186
#Not Recommended	3,050	61,024	117,353	99,912
#Total	27,956	458,114	307,129	826,098

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL DATASET

In this study, we opted to use the Steam platform, which offers a wide range of games and comprehensive reviews and discussions. Our selection of four representative esports games - TEKKEN 7 (Fighting), Dota2 (MOBA), PUBG (BR), and CS:GO (FPS) - was informed by survey reports from Newzoo² and existing esports events. We used Steam’s API³ to crawl game reviews and gather a new dataset of English reviews for the four chosen esports games up to December 2021. The dataset features unique identifiers, such as *recommendation ID, user ID, review publication and updated date, playtime when the review was written, recommend/not recommend tag, language tag, review content, and the number of people who rated the review as helpful, funny, and more.*

To prepare the data for LDA topic modeling, we followed a standard text preprocessing approach that consisted of three primary phases: (1) tokenization, (2) normalization, and (3) noise removal. This involved sequentially removing HTML tags and whitespaces, deleting special characters, converting all texts to lowercase, and eliminating stop words based on the English stop word list. Finally, we used lemmatization to convert words into their base form, such as “wolves” to “wolf” or “eating” to “eat”. This process resulted in a refined dataset for topic modeling. Table 2 shows the number of positive, negative, and total reviews for each esports game after preprocessing.

²<https://newzoo.com/consumer-insights/games-esports>

³<https://partner.steamgames.com/doc/store/getreviews>

TABLE 3. An annotated sentence with labels in BIEOS encoding.

Words:	Absolute	trash	.	Toxic	community	and	smurfs	with	a	trashy	anti	cheat	system	.
Labels:	O	O	O	O	S-NEG	O	S-NEG	O	O	O	B-NEG	I-NEG	E-NEG	O

Games are long-term services that are greatly influenced by the time players engage in them. Lin et al. [7] conducted a study that found that game feedback heavily depends on total playtime, especially in the esports industry, where games are the main focus. Fig. 3 illustrates the distribution of positive and negative reviews based on the playtime. The vertical axis represents playtime in hours, while the horizontal axis shows the positive and negative distributions for the four games. Generally, players in the early stages of playing a game tend to provide more negative reviews than positive ones. However, as they spent more time playing, their reviews became less negative and more positive. This is reasonable because players may not be familiar with graphics, stories, maps, and characters at the beginning, but their feedback becomes more positive as they become more experienced. Additionally, game producers may improve game services based on player feedback; therefore, the longer a player engages in a game, the better the service is likely to become.

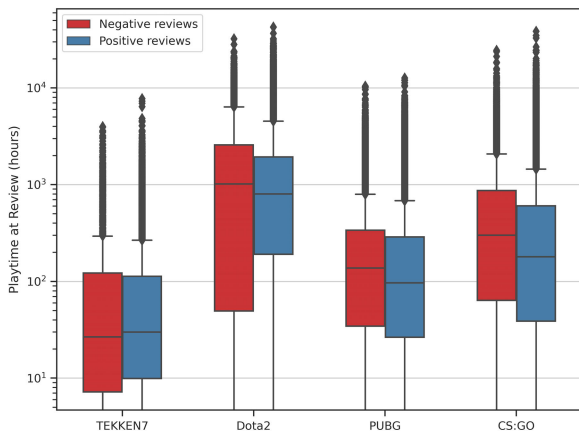


FIGURE 3. The positive/negative distribution of four esports reviews.

B. ANNOTATED ESPORTS DATASET

We developed a BERT-based model for the ABSA task by creating a new annotated dataset from esports game reviews. We manually extracted 3,100 sentences and invited three equally experienced annotators to provide aspect-level annotations. To ensure accuracy, we assigned each annotator 1,000 sentences from four esports games and 100 test sentences to measure the inter-annotator agreement.

The annotators focused solely on explicit aspects with polarities mentioned in the sentences, using a BIEOS encoding format that included the tags B-{POS, NEU, NEG}, I-{POS, NEU, NEG}, E-{POS, NEU, NEG}, O-{POS, NEU, NEG}, and S-{POS, NEU, NEG}. These tags represent the Beginning, Inside, End, and Outside of an aspect,

TABLE 4. Inter-annotator agreement for three annotators.

Annotator	TEKKEN7	Dota2	PUBG	CS:GO
A1 & A2	0.8811	0.8784	0.9087	0.9309
A2 & A3	0.8636	0.8669	0.8932	0.8840
A1 & A3	0.8522	0.8365	0.8741	0.8512
#Average	0.8657	0.8606	0.8920	0.8887

respectively, where S indicates a single word as an aspect. POS, NEU, and NEG denote the Positive, Neutral, and Negative polarities, respectively. An example of an annotated sentence is presented in Table 3.

We calculated inter-annotator agreement (IAA) using 100 test sentences to assess the consistency rate among the annotators. In this study, we evaluated IAA using Cohen’s kappa (κ) [42]. The formula for Cohen’s kappa is as follows:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{7}$$

Here, p_o reflects the relative observed agreement among the annotators, and p_e represents the hypothetical probability of chance agreement. It’s worth noting that Cohen’s kappa is designed for use with two annotators; therefore, we calculated it for each pair of our annotators. The pairwise agreement was calculated for each pair of annotators, and the average values are presented in Table 4. Based on our results, it appears that the three annotators strongly agree with each other. This conclusion is supported by the high kappa value obtained, as defined by Viera and Garrett’s (2005) scale [43], where a kappa of 1 indicates perfect agreement and a kappa of 0 indicates agreement equivalent to chance.

C. TOPIC MODELING AND GROUPING WITH LDA

We employed the LDA algorithm to uncover the topics and keywords represented in four esports games. We used the coherence value [44] to determine the optimal number of topics. Fig. 4 shows the coherence values for topic numbers ranging from 3 to 100. We chose 16 topics each for TEKKEN7, Dota2, and CS:GO, while selecting 15 topics for PUBG, as the coherence values flattened out at these optimal values.

To provide insight into the topics identified through LDA, we present the representative topics from each esports game using WordCloud in Fig. 5. In each WordCloud, the size of each word corresponds to its probability for one topic.

To validate the topics inferred by the LDA model, we combined the reading of topic words from WordCloud with the original review text to avoid implicit expressions or forum spams. The resulting topics are listed in Table 5. These topics are classified into two groups based on their attributes.

TABLE 5. Keyword examples and inferred topics of esports game.

(a) TEKKEN7

ID	Inferred topics	Keyword examples (TEKKEN7)
1	graphics	graphic*0.002, physics*0.002
2	character	character*0.041, customization*0.001
3	map	street*0.001, stage*0.002
4	optimization	bug*0.002, issue*0.001
5	update	dlc*0.002, season*0.009
6	gameplay	mode*0.021, arcade*0.004, multiplayer*0.001
7	community	community*0.001, fan*0.002, people*0.009
8	server	server*0.001, connection*0.004
9	region	korean*0.001, japan*0.001
10	price	money*0.002, price*0.001
11	matchmaking	match*0.001, matchmaking*0.001
12	cheating	smurf*0.001, denuvo*0.001
13	skill	combo*0.010, punch*0.005, kick*0.003
14	learning curve	learning*0.001, curve*0.001, tutorial*0.001
15	ranking	rank*0.005, leaderboard*0.001
16	story	story*0.017, content*0.002

(b) Dota2

ID	Inferred topics	Keyword examples (Dota2)
1	graphics	graphic*0.007, fps*0.002
2	character	hero*0.022, character*0.003
3	map	map*0.002, rock*0.002
4	optimization	bug*0.001, issue*0.001
5	update	update*0.033, version*0.003
6	gameplay	moba*0.065, multiplayer*0.005
7	community	community*0.048, people*0.038
8	server	server*0.024
9	region	SEA*0.019, language*0.007
10	price	money*0.003
11	matchmaking	match*0.013, matchmaking*0.004
12	cheating	smurf*0.004
13	skill	farm*0.007, lifesteal*0.002
14	learning curve	learning*0.010, curve*0.012, noob*0.015
15	teamwork	team*0.016, teammate*0.005
16	ranking	rank*0.001, mmr*0.012

(c) PUBG

ID	Inferred topics	Keyword examples (PUBG)
1	graphics	graphic*0.009, fps*0.004
2	character	player*0.018, skin*0.004
3	map	map*0.012, zone*0.001
4	optimization	crash*0.020, optimization*0.003, bug*0.025
5	teamwork	team*0.001, friend*0.015
5	update	update*0.011, release*0.003
6	gameplay	gameplay*0.006, battle*0.010, mode*0.001
7	community	people*0.013, relationship*0.003
8	server	server*0.022, lag*0.006
9	region	region*0.021, china*0.024
10	price	money*0.012, loot*0.002, price*0.002
11	matchmaking	match*0.004, hour*0.003
12	cheating	hacker*0.022, cheater*0.036
13	skill	gun*0.003, shoot*0.001
14	learning curve	learning*0.001, curve*0.001
15	teamwork	team*0.001, friend*0.015

(d) CS:GO

ID	Inferred topics	Keyword examples (CS:GO)
1	graphics	graphic*0.008, fps*0.009
2	character	terrorist*0.008, character*0.002
3	map	map*0.008
4	optimization	problem*0.008, issue*0.001, fix*0.001
5	update	update*0.001, version*0.001
6	gameplay	mode*0.001, multiplayer*0.005
7	community	community*0.035, fan*0.002, people*0.049
8	server	server*0.012, account*0.001
9	region	russian*0.039, language*0.001
10	price	money*0.007, price*0.005
11	matchmaking	match*0.001, matchmaking*0.001
12	cheating	hacker*0.121, cheater*0.043, ban*0.002
13	skill	bomb*0.007, nuke*0.006, fire*0.006
14	learning curve	learning*0.003, curve*0.004, noob*0.008
15	teamwork	team*0.012, friend*0.046, mate*0.003
16	ranking	rank*0.002, silver*0.007

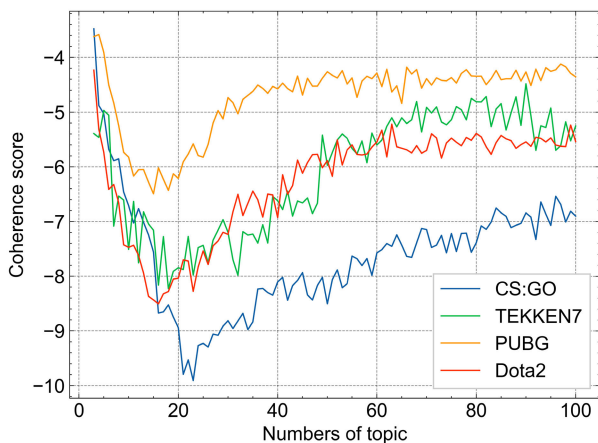


FIGURE 4. Coherence scores of four esports games.

a: GAME-RELATED TOPICS (GRT)

In Table 5, the first six inferred topics (ID=1-6) represent common topics mentioned in all four esports games,



FIGURE 5. Topics discovered by LDA.

including “graphics”, “character”, “map”, “optimization”, “update”, and “gameplay. These topics reflect fundamental and standard elements of game design. For instance, the game’s “graphics” and “gameplay” are typically determined during the game’s development, while the “character design” is often revealed to players through trailers and advertising

campaigns. Although there may be updates and optimizations after the game’s release, these factors are more related to the game operators than to the players themselves. In summary, Game-related Topics describe elements primarily driven by the operator and have little to do with player behavior.

b: PLAYER-RELATED TOPICS (PRT)

In Table 5, topics with IDs from 7 to 16 (excluding *PUBG*, which lacks the “ranking” topic) are present in all four analyzed esports games. These topics reflect the interactions between esports and their players. For instance, discussions on “skill” are mostly related to players’ understanding of the game’s mechanics, environment, or strategies. Some players may use third-party software to gain an unfair advantage, and reviews mentioning this issue are classified under the “cheating” topic. Additionally, because esports games are played online, a low-latency network is crucial for a high-quality gaming experience. Keywords under the “server” topic include lag, connection, and other related terms.

However, due to various factors and the specific meaning of each topic word being somewhat dependent on the particular esports game, the keywords and their meanings under PRT are less clear than those under GRT.

D. TOPIC PREVALENCE ANALYSIS RESULTS

In this section, we present an analysis of topic prevalence using the LDA model to investigate the influence of various aspects on esports players. Fig. 6 illustrates the interaction between esports players’ concerns and different esports games based on the inferred topics. The horizontal axis represents the esports game, while the vertical axis shows the inferred topic words. Each cell in the heatmap corresponds to its frequency in that game, with brighter colors indicating the higher frequency and darker colors indicating lower frequency. From the heatmap, we can observe that players’ preferences vary for different genres of esports games.

Regarding game design-related elements, traditional PC games typically focus on “graphics, narration (or story), and gameplay”. Even in the era of esports games that emphasize player versus player, the remaining two elements still play essential roles in the player’s gaming experience, excluding the story. The character is of utmost importance because of the lack of narrative. *PUBG* has been widely criticized for its optimization issues. However, the remaining three games all have character design as a top-frequency topic, indicating that a good character design can easily attract players’ attention in game design.

Furthermore, among the topic words that reflect the interaction between players and esports games, players agree that the community is crucial to their gaming experiences. A good community environment can encourage players to communicate and enhance their gaming experience. Different games have different features; for instance, *TEKKEN7* and *PUBG* emphasize player fighting alone, so individual player skill is significant, whereas it is less prominent in the other two games. Moreover, for *PUBG* and *CS:GO*, both shooter

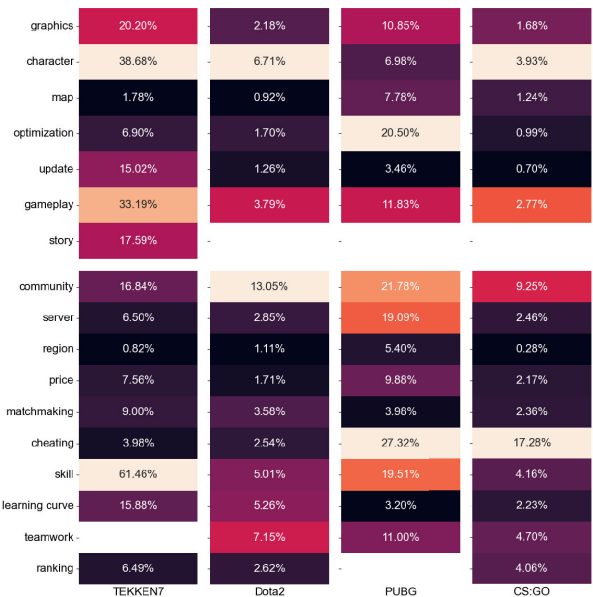


FIGURE 6. Topics prevalence among four esports games.

TABLE 6. Confusion matrix for three sentiment labels.

Confusion Matrix		Actual		
		POS	NEU	NEG
Predicted	POS	a	d	g
	NEU	b	e	h
	NEG	c	f	i

games, cheating is a major problem that needs to be addressed by operators. In summary, players have diverse concerns in different esports games that require operators’ attention and adjustment.

E. TOPIC SENTIMENT ANALYSIS

1) PERFORMANCE EVALUATION METHOD

To evaluate the performance of our sentiment classification, we employed a commonly used three-class confusion matrix, as presented in Table 6, to describe the classification results. The data prediction results can yield one of four possible outcomes: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). In this context, True (T) and False (F) indicate whether the predicted result is correct or incorrect, while Positive (P) and Negative (N) describe the predicted samples as having positive or negative sentiment, respectively.

Our model utilized three sentiment labels, represented as $T = POS, NEU, NEG$ for Positive, Neutral, and Negative sentiments. Equations 8, 9, and 10 define the formulas for calculating the TP, FP, TN, and FN for each label.

- For sentiment label is POS:

$$\begin{aligned}
 TP_{POS} &= a & TN_{POS} &= e + f + h + i \\
 FP_{POS} &= d + g & FN_{POS} &= b + c
 \end{aligned} \tag{8}$$

- For sentiment label is NEU:

$$\begin{aligned} TP_{NEU} &= e & TN_{NEU} &= a + c + g + i \\ FP_{NEU} &= b + h & FN_{NEU} &= d + f \end{aligned} \quad (9)$$

- For sentiment label is NEG:

$$\begin{aligned} TP_{NEG} &= i & TN_{NEG} &= a + b + d + e \\ FP_{NEG} &= c + f & FN_{NEG} &= g + h \end{aligned} \quad (10)$$

a: MACRO PERFORMANCE

To obtain *Precision (Pre)*, *Recall*, and *Macro F₁* score under macro criteria, we first calculated the precision and recall for each sentiment label *t* as follows:

$$Pre_t = \frac{TP_t}{TP_t + FP_t} \quad Recall_t = \frac{TP_t}{TP_t + FN_t} \quad (11)$$

And the *F₁* score under label *t* is defined as the following Equation:

$$F_{1t} = \frac{2 * Pre_t * Recall_t}{Pre_t + Recall_t} \quad (12)$$

Thus, the *Macro F₁* is defined as:

$$Macro F_1 = \frac{\sum_{t \in T} F_{1t}}{3} \quad (13)$$

b: MICRO PERFORMANCE

For *Micro F₁*, we first counted the total TP, FN, and FP among all sentiment labels to obtain the global *Pre* and *Recall*:

$$\begin{aligned} Pre_{mi} &= \frac{\sum_{t \in T} TP_t}{\sum_{t \in T} TP_t + \sum_{t \in T} FP_t} \\ Recall_{mi} &= \frac{\sum_{t \in T} TP_t}{\sum_{t \in T} TP_t + \sum_{t \in T} FN_t} \end{aligned} \quad (14)$$

Similar to *Macro F₁*, we calculated *Micro F₁* as follows:

$$Micro F_1 = \frac{2 * Pre_{mi} * Recall_{mi}}{Pre_{mi} + Recall_{mi}} \quad (15)$$

2) MODEL SETTINGS AND PERFORMANCE

In this study, we used the pre-trained “bert-base-uncased” model with default parameter settings and single-layer architecture as the downstream component for E2E-ABSA. We divided the overlength reviews into paragraphs to ensure that the input token length remained within the maximum sequence length limit. We trained the combined models for up to 1,500 steps and selected the best model based on the Micro *F₁* and Macro *F₁* scores of the test set. To increase the reliability of our results, we repeated the training process four times with different random seeds, resulting in a total of 30 models, and reported the average results. Our experimental results showed that the BERT-TFM model achieved Micro *F₁* and Macro *F₁* scores of 75.48% and 60.46%, respectively. Therefore, we utilized the BERT-TFM model to perform the E2E-ABSA task on esports game reviews.

3) SENTIMENT CLASSIFICATION AND SUMMARIZATION

We investigated the trend of negative reviews as a percentage of the total reviews by examining the distribution of negative reviews over playtime, as shown in Fig. 7. Our findings build upon previous research and are consistent with the observations made in Section IV-A. The results reveal that all games have a higher percentage of negative reviews when playtime is less than 2 hours.

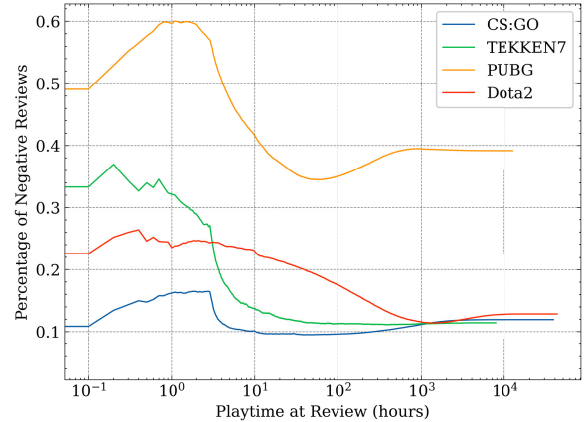


FIGURE 7. Distribution of Negative Reviews by Playtime at Review (hours).

To gain a more accurate understanding of the sentiments expressed by novice players, we specifically extracted reviews with a playtime of less than 10 hours. Fig. 8 presents the sentiment classification results of our proposed E2E-ABSA framework for the four esports games, with positive, neutral, and negative sentiment rates represented by green, yellow, and red, respectively. For novice players with playtime of less than 10 hours, we use lighter shades of these colors to indicate their sentiments.

Regarding the design of traditional game elements (GRT group), all four games demonstrate strong performance in terms of “graphics” and “gameplay”, with more than half of the feedback was positive. However, apart from TEKKEN7, the other three esports games struggle with “optimization” issues, particularly PUBG. Most players agree that PUBG faces significant optimization issues that overshadow its otherwise excellent performance in areas such as “map” and “character”. Furthermore, players in all four games reported dissatisfaction with “updates”. As the only way to fix bugs and add new content to the game, successful updates significantly enhance players’ gaming experiences. For “graphics” and “gameplay”, which remain unchanged since the game’s release, it is essential to improve the quality of each game update. We believe that dissatisfaction with updates indirectly reduces players’ evaluations of optimization. Overall, the four esports games received good ratings in the basic game design elements, but follow-up services, such as updates, were relatively unsatisfactory.

Cheating is an intolerable behavior in any game and can ruin the gaming experience of players, as mentioned in the reviews we examined. Therefore, effectively combating and



FIGURE 8. Topics sentiments among four esports games.

eliminating cheating is critical to the survival of any game. The second issue is matchmaking and ranking, which are especially important in Player versus Player games. Matching opponents with similar strengths is crucial to ensure a fair game, as too weak an opponent can tire players, and too strong an opponent can be frustrating, leading to a poor gaming experience. While TEKKEN7 performs relatively well in this regard, the other three games require improvement. In addition to matchmaking, stable network service is crucial for ensuring smooth gameplay. Except for TEKKEN7, the non-negative rates ($POS + NEU$) of the remaining three games under “server” are approximately lower than 20%, indicating an area where esports operators can improve. Furthermore, discussing game tips in a community composed mainly of player groups can help improve player skills and raise their competitive level, promoting the sustainable development of the gaming environment. TEKKEN7 performed the best in this regard, whereas Dota2 performed the worst. Overall, the performance of the four esports games on PRT is inferior to their performance on GRT. As esports games are a service rather than a product, follow-up service deserves as much attention from esports operators as the game’s development.

We investigated the sentiment distribution of novice players who played for less than 10 hours. Our findings show that the sentiment distribution of players’ gaming experience is not significantly different from that of the entire player base. This suggests that around 10 hours of gameplay is adequate to identify potential issues within the game. To attract and retain

more players for their esports games in the future, esports operators should prioritize enhancing the gaming experience during the initial 10 hours of gameplay. Improving the initial experience of novice players can increase their engagement and satisfaction, resulting in an expanded player base. Game developers and publishers should focus on this aspect to increase the popularity and longevity of their esports games.

F. SUMMARY

In general, this study effectively captures player preferences. By integrating the findings from prevalence and sentiment analyses, we suggest that esports operators focus on resolving issues that exhibit a high frequency in prevalence analysis while displaying low positive rates. Within the GRT group, this includes “updates” for TEKKEN 7 and “optimization” for PUBG. Meanwhile, the PRT group includes “learning curve” for TEKKEN 7; “community” and “learning curve” for Dota 2; “server”, “price”, “cheating”, and “teamwork” for PUBG; “cheating” and “ranking” for CS:GO.

It is important to note that the remaining topics highlight other significant concerns. However, when esports operators prioritize the aforementioned problems while working with limited resources, it will result in more immediate enhancements of the gaming experience. Building on the foundations of this study and considering the elements of time, our proposed model can be used to track player feedback. Consequently, this will enable esports operators to engage with players more promptly and contribute to an improved gaming experience.

Through the analysis, we also explored the niche status of the Fighting game genre and the disparities within the TEKKEN series. In TEKKEN 7, removing Team Battle resulted in significantly fewer reviews for that game than for the other three. In future research, gathering more data on Fighting Games is essential to better understand the player requirements.

V. CONCLUSION

In esports, unbalanced gaming experience between professional players and noobs is a problem caused by the lack of methods for operators to quickly obtain useful information from noob players' feedback. To address this, we propose a hybrid approach of topic modeling and sentiment analysis to automatically analyze the vast number of game reviews from noob players. This will enable esports operators to better target their opinions and build a more balanced gaming environment.

Our analysis of four representative esports games, TEKKEN7, Dota2, PUBG, and CS:GO, yielded several important insights. We extracted and summarized 16, 16, 15, and 16 topics for each game and divided them into GRT and PRT topics. We found that players value graphics, gameplay, and character in GRT, while their preferred PRT topics depend on the specific game, such as skill for TEKKEN7 and cheating for PUBG and CS:GO.

We also employed a BERT-based model for sentiment analysis using the E2E-ABSA task on experimental datasets that reflected players' attitudes toward different topics. Our analysis indicated that negative attitudes toward PRT topics were generally higher than those toward GRT topics, suggesting that esports game operators should shift from being product providers to service providers.

Furthermore, our examination of reviews from novice players with less than 10 hours of playtime revealed additional feedback. This emphasizes that esports operators should prioritize enhancing the gaming experience during the first 10 hours. By doing so, they can increase player engagement and satisfaction, leading to greater retention and expansion of the player base.

This study had some limitations. First, we focused only on analyzing English-language reviews and did not include all popular games, such as the League of Legends in MOBAs. Given the importance of the Asian market, we plan to expand our approach to include non-English reviews in the future. Second, we analyzed only player reviews on Steam. Thus, we plan to incorporate other platforms, such as Metacritic, in our future research to obtain a more comprehensive understanding of the sentiment of esports game players. Finally, we aim to investigate game cultures and hotspots to extract more valuable information, which can be used to enhance the services provided by esports operators.

REFERENCES

[1] 2022 Asian Games. Accessed: Sep. 2022. [Online]. Available: <https://www.aesf.com/en/News-And-Media-Coverage/Esports>Returns-As-Proposed-By-The-Hagoc-For-The-2022-Asian-Games.html>

[2] *Global Esports Market in 2020*. Accessed: Sep. 2022. [Online]. Available: <https://newzoo.com/insights/trend-reports/newzoo-global-esports-market-report-2020-light-version/>

[3] F. Baumann, D. Emmert, H. Baumgartl, and R. Buettner, "Hardcore gamer profiling: Results from an unsupervised learning approach to playing behavior on the steam platform," *Proc. Comput. Sci.*, vol. 126, pp. 1289–1297, Sep. 2018, doi: [10.1016/j.procs.2018.08.078](https://doi.org/10.1016/j.procs.2018.08.078).

[4] A. Agrawal, C. Silva, and P. Xenopoulos, "Data science and analytics for esports," in *Proc. IEEE Integr. STEM Educ. Conf. (ISEC)*, Mar. 2022, p. 386, doi: [10.1109/ISEC54952.2022.10025321](https://doi.org/10.1109/ISEC54952.2022.10025321).

[5] A. Wattanaburanon and N. Propoon, "Method for classifying usability qualities and problems for action games from user reviews using text mining," in *Proc. IEEE/ACIS 15th Int. Conf. Comput. Inf. Sci. (ICIS)*, Jun. 2016, pp. 1–6, doi: [10.1109/ICIS.2016.7550748](https://doi.org/10.1109/ICIS.2016.7550748).

[6] N. H. L. Wong, S.-I. Kwok, and J. L. Tan, "Predictive analytics of first blood and match outcome in Dota 2," in *Proc. IEEE Region 10 Conf. (TENCON)*, Nov. 2022, pp. 1–6, doi: [10.1109/TENCON55691.2022.9977786](https://doi.org/10.1109/TENCON55691.2022.9977786).

[7] D. Lin, C.-P. Bezemer, Y. Zou, and A. E. Hassan, "An empirical study of game reviews on the steam platform," *Empirical Softw. Eng.*, vol. 24, no. 1, pp. 170–207, Feb. 2019, doi: [10.1007/s10664-018-9627-4](https://doi.org/10.1007/s10664-018-9627-4).

[8] H.-N. Kang, "A study of analyzing on online game reviews using a data mining approach: STEAM community data," *Int. J. Innov. Manage. Technol.*, vol. 8, no. 2, pp. 90–94, 2017, doi: [10.18178/ijimt.2017.8.2.709](https://doi.org/10.18178/ijimt.2017.8.2.709).

[9] B. Liu, *Sentiment Analysis and Opinion Mining*. San Rafael, CA, USA: Morgan & Claypool, 2012, doi: [10.1007/978-3-031-02145-9](https://doi.org/10.1007/978-3-031-02145-9).

[10] F. Tang, L. Fu, B. Yao, and W. Xu, "Aspect based fine-grained sentiment analysis for online reviews," *Inf. Sci.*, vol. 488, pp. 190–204, Jul. 2019, doi: [10.1016/j.ins.2019.02.064](https://doi.org/10.1016/j.ins.2019.02.064).

[11] C. R. Aydın and T. Güngör, "Combination of recursive and recurrent neural networks for aspect-based sentiment analysis using inter-aspect relations," *IEEE Access*, vol. 8, pp. 77820–77832, 2020, doi: [10.1109/ACCESS.2020.2990306](https://doi.org/10.1109/ACCESS.2020.2990306).

[12] N. Li, C.-Y. Chow, and J.-D. Zhang, "SEML: A semi-supervised multi-task learning framework for aspect-based sentiment analysis," *IEEE Access*, vol. 8, pp. 189287–189297, 2020, doi: [10.1109/ACCESS.2020.3031665](https://doi.org/10.1109/ACCESS.2020.3031665).

[13] Z. Kastrati, A. S. Imran, and A. Kurti, "Weakly supervised framework for aspect-based sentiment analysis on students' reviews of MOOCs," *IEEE Access*, vol. 8, pp. 106799–106810, 2020, doi: [10.1109/ACCESS.2020.3000739](https://doi.org/10.1109/ACCESS.2020.3000739).

[14] W. Li and B. Xu, "Aspect-based fashion recommendation with attention mechanism," *IEEE Access*, vol. 8, pp. 141814–141823, 2020, doi: [10.1109/ACCESS.2020.3013639](https://doi.org/10.1109/ACCESS.2020.3013639).

[15] J. H. Hyeong, K. J. Choi, J. Y. Lee, and T.-H. Pyo, "For whom does a game update? Players' status-contingent gameplay on online games before and after an update," *Decis. Support Syst.*, vol. 139, Dec. 2020, Art. no. 113423, doi: [10.1016/j.dss.2020.113423](https://doi.org/10.1016/j.dss.2020.113423).

[16] E. Conroy, M. Kowal, A. J. Toth, and M. J. Campbell, "Boosting: Rank and skill deception in esports," *Entertainment Comput.*, vol. 36, Jan. 2021, Art. no. 100393, doi: [10.1016/j.entcom.2020.100393](https://doi.org/10.1016/j.entcom.2020.100393).

[17] J. P. Zagal, N. Tomuro, and A. Shepitsen, "Natural language processing in game studies research: An overview," *Simul. Gaming*, vol. 43, no. 3, pp. 356–373, 2012, doi: [10.1177/1046878111422560](https://doi.org/10.1177/1046878111422560).

[18] B. Strååt, H. Verhagen, and H. Warpefelt, "Probing user opinions in an indirect way: An aspect based sentiment analysis of game reviews," in *Proc. 21st Int. Academic Mindtrek Conf.*, Sep. 2017, pp. 1–7, doi: [10.1145/3131085.3131121](https://doi.org/10.1145/3131085.3131121).

[19] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. AL-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. M. Jiménez-Zafra, and G. Eryigit, "SemEval-2016 task 5: Aspect based sentiment analysis," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 19–30, doi: [10.18653/v1/S16-1002](https://doi.org/10.18653/v1/S16-1002).

[20] K. Yauris and M. L. Khodra, "Aspect-based summarization for game review using double propagation," in *Proc. Int. Conf. Adv. Informat., Concepts, Theory, Appl. (ICAICTA)*, Aug. 2017, pp. 1–6, doi: [10.1109/ICAICTA.2017.8090997](https://doi.org/10.1109/ICAICTA.2017.8090997).

[21] B. Strååt and H. Verhagen, "Using user created game reviews for sentiment analysis: A method for researching user attitudes," in *Proc. GHITALLY@CHIItaly*, Apr. 2017, pp. 1–6.

[22] M. K. Baowaly, Y.-P. Tu, and K.-T. Chen, "Predicting the helpfulness of game reviews: A case study on the steam store," *J. Intell. Fuzzy Syst.*, vol. 36, no. 5, pp. 4731–4742, May 2019, doi: [10.3233/JIFS-179022](https://doi.org/10.3233/JIFS-179022).

- [23] S. Ruseti, M.-D. Sirbu, M. A. Calin, M. Dascalu, S. Trausan-Matu, and G. Militaru, "Comprehensive exploration of game reviews extraction and opinion mining using NLP techniques," in *Proc. 4th Int. Congr. Inf. Commun. Technol.* Singapore: Springer, 2020, pp. 323–331, doi: [10.1007/978-981-15-0637-6_27](https://doi.org/10.1007/978-981-15-0637-6_27).
- [24] M. Hu, Y. Peng, Z. Huang, D. Li, and Y. Lv, "Open-domain targeted sentiment analysis via span-based extraction and classification," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2019, pp. 537–546, doi: [10.18653/v1/P19-1051](https://doi.org/10.18653/v1/P19-1051).
- [25] X. Ma and E. Hovy, "End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics, Long Papers*, vol. 1. Berlin, Germany: Association for Computational Linguistics, 2016, pp. 1064–1074, doi: [10.18653/v1/P16-1101](https://doi.org/10.18653/v1/P16-1101).
- [26] M. Schmitt, S. Steinheber, K. Schreiber, and B. Roth, "Joint aspect and polarity classification for aspect-based sentiment analysis with end-to-end neural networks," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2018, pp. 1109–1114, doi: [10.18653/v1/D18-1139](https://doi.org/10.18653/v1/D18-1139).
- [27] X. Li, L. Bing, W. Zhang, and W. Lam, "Exploiting BERT for end-to-end aspect-based sentiment analysis," in *Proc. 5th Workshop Noisy User-Generated Text (W-NUT)*, 2019, pp. 34–41, doi: [10.18653/v1/D19-5505](https://doi.org/10.18653/v1/D19-5505).
- [28] A. Agarwal and J. S. Sabharwal, "End-to-end sentiment analysis of Twitter data," in *Proc. Workshop Inf. Extraction Entity Anal. Social Media Data*, 2012, pp. 39–44.
- [29] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Jan. 2003, doi: [10.1162/jmlr.2003.3.4-5.993](https://doi.org/10.1162/jmlr.2003.3.4-5.993).
- [30] J. Huang, S. Rogers, and E. Joo, "Improving restaurants by extracting subtopics from yelp reviews," in *Proc. iConf. Social Media Expo*, 2014, pp. 1–5. [Online]. Available: <https://hdl.handle.net/2142/48832>
- [31] S. Feuerriegel, A. Ratku, and D. Neumann, "Analysis of how underlying topics in financial news affect stock prices using latent Dirichlet allocation," in *Proc. 49th Hawaii Int. Conf. Syst. Sci. (HICSS)*, Jan. 2016, pp. 1072–1081, doi: [10.1109/HICSS.2016.137](https://doi.org/10.1109/HICSS.2016.137).
- [32] T. Santos, F. Lemmerich, M. Strohmaier, and D. Helic, "What's in a review: Discrepancies between expert and amateur reviews of video games on metacritic," in *Proc. ACM Hum.-Comput. Interact.*, 2019, pp. 1–22, doi: [10.1145/3359242](https://doi.org/10.1145/3359242).
- [33] Y. Heng, Z. Gao, Y. Jiang, and X. Chen, "Exploring hidden factors behind online food shopping from Amazon reviews: A topic mining approach," *J. Retailing Consum. Services*, vol. 42, pp. 161–168, May 2018, doi: [10.1016/j.jretconser.2018.02.006](https://doi.org/10.1016/j.jretconser.2018.02.006).
- [34] I. Putri and R. Kusumaningrum, "Latent Dirichlet allocation (LDA) for sentiment analysis toward tourism review in Indonesia," *J. Phys., Conf. Ser.*, vol. 801, Jan. 2017, Art. no. 012073, doi: [10.1088/1742-6596/801/1/012073](https://doi.org/10.1088/1742-6596/801/1/012073).
- [35] T. Tran, H. Ba, and V.-N. Huynh, "Measuring hotel review sentiment: An aspect-based sentiment analysis approach," in *Integrated Uncertainty in Knowledge Modelling and Decision Making*. Cham, Switzerland: Springer, 2019, pp. 393–405, doi: [10.1007/978-3-030-14815-7_33](https://doi.org/10.1007/978-3-030-14815-7_33).
- [36] B. Gifford, "Reviewing the critics: Examining popular video game reviews through a comparative content analysis," Ph.D. thesis, School Commun., Cleveland State Univ., Cleveland, OH, USA, 2013.
- [37] M. Bond and R. Beale, "What makes a good game? Using reviews to inform design," in *Proc. People Comput. XXIII Celebrating People Technol.*, 2009, pp. 418–422.
- [38] I. J. Livingston, L. E. Nacke, and R. L. Mandryk, "The impact of negative game reviews and user comments on player experience," in *Proc. ACM SIGGRAPH Game Papers*, Aug. 2011, pp. 25–29, doi: [10.1145/2037692.2037697](https://doi.org/10.1145/2037692.2037697).
- [39] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.* Stroudsburg, PA, USA: Association for Computational Linguistics, 2019, pp. 4171–4186, doi: [10.18653/v1/N19-1423](https://doi.org/10.18653/v1/N19-1423).
- [40] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1–11.
- [41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [42] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960, doi: [10.1177/001316446002000104](https://doi.org/10.1177/001316446002000104).
- [43] A. J. Viera and J. M. Garrett, "Understanding interobserver agreement: The kappa statistic," *Family Med.*, vol. 37, no. 5, pp. 360–363, 2005. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/15883903/>
- [44] D. Newman, J. H. Lau, K. Grieser, and T. Baldwin, "Automatic evaluation of topic coherence," in *Proc. Hum. Lang. Technol., Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics*, 2010, pp. 100–108, doi: [10.5555/1857999.1858011](https://doi.org/10.5555/1857999.1858011).



YANG YU received the master's degree in knowledge science from the Japan Advanced Institute of Science and Technology (JAIST), in 2020, where he is currently pursuing the Ph.D. degree in artificial intelligence. His research interests center around the sentiment analysis of natural language processing. His current research interests include topic mining and sentiment analysis research into the game reviews from game platforms and communities.



DUY-TAI DINH received the Ph.D. degree from the Japan Advanced Institute of Science and Technology (JAIST), in 2020. He is currently an Assistant Professor with The Kyoto College of Graduate Studies for Informatics, Japan. His research interest includes the various domains of data science, including data mining, machine learning, and materials informatics. He serves as an Editorial Board Member for the *Journal of Computer Science and Data Science and Management* journal.



BA-HUNG NGUYEN received the Ph.D. degree from the Japan Advanced Institute of Science and Technology (JAIST), in 2020. He was a Post-graduate Researcher with the Management Science and Business Economics (MSBE) Group, Business School, The University of Edinburgh, U.K. He is currently the Dean of the Faculty of Economics and Management, Thai Binh Duong University, Vietnam. His research interest includes AI applications in business and management and data-driven insights for decision-making systems.



FANGYU YU received the Diploma degree in health sciences (nursing) from Ngee Ann Polytechnic, Singapore, and the master's degree in knowledge science from the Japan Advanced Institute of Science and Technology (JAIST), in 2023. Her research interests include deep neural networks, web languages, and data analysis. Her greatest passion is coding into a real-life project and using her technical know-how to benefit other people and organizations.



VAN-NAM HUYNH (Member, IEEE) received the Ph.D. degree in mathematics from the Vietnam Academy of Science and Technology, in 1999. He is currently a Professor with the School of Knowledge Science, Japan Advanced Institute of Science and Technology (JAIST). His current research interests include machine learning and data mining, AI reasoning, argumentation, multi-agent systems, decision analysis, management science, and Kansei information processing and applications. He currently serves as an Area Editor for *International Journal of Approximate Reasoning*, the Editor-in-Chief for *International Journal of Knowledge and Systems Science*, and the Editorial Board Member of the *Array* journal.