engineering proceedings

Proceeding Paper

Using a Text Mining Approach to Identify Important Factors Influencing the Performance of Programmatic Advertising †

Yi-Yun Wang, Venkateswarlu Nalluri and Long-Sheng Chen *

Department of Information Management, Chaoyang University of Technology, Taichung 413310, Taiwan; winona25544@gmail.com (Y.-Y.W.); nallurivenkey7@gmail.com (V.N.)

* Correspondence: lschen@cyut.edu.tw

† Presented at the 3rd IEEE International Conference on Electronic Communications, Internet of Things and Big Data Conference 2023, Taichung, Taiwan, 14–16 April 2023.

Abstract: Programmatic advertising uses big data to spread personalized marketing materials to target audiences, which is a major driving force for the growth of digital advertising. Among them, in-application advertisements (in-app ads) are an important part of programmatic advertising. In in-app advertising, which is highly related to application revenue, ads are delivered to customers through mobile devices at any time and place based on personal needs. Due to the power of electronic Word-of-Mouth (e-WOM), text comments from social media are becoming a new mode of advertising, influencing consumers' purchase behavior. Text reviews on social media are more powerful than traditional ads. However, relatively little research has studied this issue. Therefore, using text mining and latent semantic analysis techniques, we aimed to discover the advertising elements of text reviews in the social community. Based on the results, suggestions were made to advertising companies to improve the performance of text reviews when employing key opinion leaders (KOL) to write commercial comments that promote products or services.

Keywords: text mining; programmatic advertising; in-app advertising; text mining; latent semantic analysis

Citation: Wang, Y.-Y.; Nalluri, V.; Chen, L.-S. Using a Text Mining Approach to Identify Important Factors Influencing the Performance of Programmatic Advertising. *Eng. Proc.* **2023**, *38*, 15. https://doi.org/10.3390/engproc2023038015

Academic Editors: Teen-Hang Meen, Hsin-Hung Lin and Cheng-Fu Yang

Published: 20 June 2023

1. Introduction

As technology continues to advance, programmatic advertising is an emerging and rapidly evolving information technology with cost-effectiveness that uses big data to deliver personalized marketing topics to target audiences [1]. In-app advertising (also known as mobile advertising) is a type of advertising that is displayed using applications (apps). In recent years, with the popularity of mobile devices, apps have been growing rapidly, causing marketers to pay more attention to the promotion and marketing capabilities of apps. Manufacturers promote their products and expand their user base through various advertising channels, and the medium of advertising has crossed over from traditional website browsing and social networking to applications [2]. According to e-Marketer in 2022, programmatic advertising is becoming an increasingly important part of the advertising landscape [3]. As a result, in-app advertising is gaining more attention.

Although programmatic advertising has become common in advertising technology, the rapid development of programmatic advertising and the immediate exposure of commercial content has fundamentally changed the traditional marketing model [4]. Many advertisers currently use programmatic advertising techniques, but there is a risk that ads may be presented on non-quality sites. Reference [1] suggested that when premium brand advertisements appeared on non-premium websites, users' negative perceptions of the advertisements and brands increased. Therefore, it has become an important issue to understand the key factors of in-app advertising in user experience to enhance the effectiveness of advertising [5].

In addition, the importance of electronic Word-of-Mouth (e-WOM) in e-commerce is gradually increasing. Consumers make purchase decisions by reading reviews. Therefore, marketers can use e-WOM to build product awareness, increase sales, enhance brand value, and build customer loyalty [6]. Many brands use high prices to purchase reviews to increase brand awareness [7,8]. Therefore, text comments from online community users are becoming a new mode of alternative advertising communication. Reference [8] also confirmed that the advertising behavior of purchasing reviews does influence consumers' perceptions.

Previous studies related to mobile advertising used questionnaires to examine the effect of personalization, entertainment, trustworthiness, and information content on users' attitudes toward receiving mobile advertising [2]. However, questionnaire methods are prone to problems such as sampling errors and do not provide a timely understanding of consumer perceptions [9]. Consequently, we used text mining and latent semantic analysis (LSA) techniques to identify the elements of text reviews. The used methodology can improve traditional questionnaire survey methods and also address the limitation of the human inability to analyze a large number of text reviews.

Due to the power of electronic e-WOM, text comments from social media have become a new mode of ads. Text reviews on social media are more powerful than traditional ads. However, relatively little works have studied this issue. Therefore, using text mining and latent semantic analysis techniques, we discovered the advertising elements of text reviews in the social community. Based on the results, we provide advertising companies with suggestions to improve the performance of text reviews when employing key opinion leaders (KOL) to write commercial comments to promote products or services.

2. Literature Review

2.1. Programmatic Advertising

Programmatic advertising is a type of marketing behavior that uses big data and artificial intelligence to precisely target advertisements to a highly personalized target audience [10]. A 10% increase in the number of programmatic ads and a 24% increase in the average price per ad were reported in a 2021 annual report [11], which shows the importance of programmatic advertising.

Reference [10] argued that programmatic advertising was a new and lesser-known advertising technology that delivered advertising messages to target audiences in real-time via the internet. Conversely, programmatic advertising may pose risks [4], focusing on programmatic advertising for alcohol. Programmatic advertising changes the original marketing paradigm beyond the regulation of commercial content exposure, undermines public regulation of alcohol marketing, and presents a new challenge to any model of public marketing oversight [4]. With the high use of programmatic advertising behavior, advertisers are also presenting their ads in various forms. Thus, companies are paying more attention to programmatic advertising and analyzing the performance of programmatic advertising is one of the important issues.

2.2. In-App Advertising

Mobile devices offer a wide variety of apps, with global app installs in the App Store and Google Play reaching nearly 37 billion in the first quarter of 2022 [12], and the ubiquity of mobile phones allows marketers to advertise regardless of time and geography [2]. Therefore, the revenue source of apps is more from in-app advertising than from sales [13], and the advertising medium ranges from traditional website browsing, social networking, and even cross-platform apps.

2.3. Text Mining

Text mining is the process of organizing and analyzing large amounts of textual data to find useful information that can be used by specific users to make decisions. There are many studies on text mining techniques for online reviews. Reference [14] used text mining to analyze feedback from online shopping reviews. Reference [15] proposed an easy-to-implement online review text analysis procedure through text mining for studying brand image and brand positioning. Reference [16] also used text exploration to build four predictors to analyze the interrelationships between reviewer credibility, review age, and review variance. Most of the existing in-app advertising studies adopted questionnaires and interviews. However, this approach requires a lot of time and capacity, and it is not possible to obtain immediate responses from users [9]. Thus, we used text mining instead of traditional questionnaires to understand the important contents of a large number of reviews.

2.4. Latent Semantic Analysis (LSA)

Natural language processing (NLP) is used to automatically represent and analyze human language [17]. Latent semantic analysis is a natural language processing approach according to the meaning of the text context to provide concepts by extracting words from the text through dimensionality reduction techniques to enable analysis, give features, and describe key themes [18]. We referred to ref. [19] for LSA. Reference [20] performed a semantic analysis of the Bengali language and proposed a predictive model for training datasets to test and evaluate the accuracy of this model against other machine learning methods [20]. Reference [21] used LSA for online hotel reviews to understand how satisfied or dissatisfied guests feel.

3. Methodology

We used text comments on social media to determine the elements of ads and collected relevant online comments from game websites. NLP and LSA were used to discover the key elements of product reviews that affected advertising. The implemental procedure contained 5 steps as follows.

Step1: Data Collection and Preprocessing

We collected online reviews through the social center on the Steam website (https://store.steampowered.com/). Comments with less than 300 words were deleted from the collected data and non-English comments were also deleted.

Step 2: Natural Language Processing
Step 2.1: Tokenization: Textual tokenization used the Natural Language Toolkit (NLTK) of Python language.
Step 2.2: Clean data: We deleted stop words such as “the”, “and”, and other less important words in this step.
Step 2.3: Lemmatization: This step reduced complex forms of a single word to its most basic form, e.g., “ate” to “eat”.
Step 2.4: Count Word Frequency: We conducted a word frequency count and deleted words with a frequency of less than 5.
Step 2.5: Build Term Document Matrix (TDM): We used TF-IDF (term frequency-inverse document frequency) in Equation (1) to create a Term Document Matrix (TDM) for further analysis.

$$TF - IDF = TF(t_i, d_i) \times \log\left(\frac{N}{N(t_i)}\right) \tag{1}$$

Step 3: Latent Semantic Analysis (LSA)

In this study, MATLAB was used to perform Singular Value Decomposition (SVD) on word document matrices to investigate the relationship between words and their contexts.

Step 3.1: SVD

The values were brought into the SVD function in MATLAB, where the equation of SVD is shown in Figure 1.

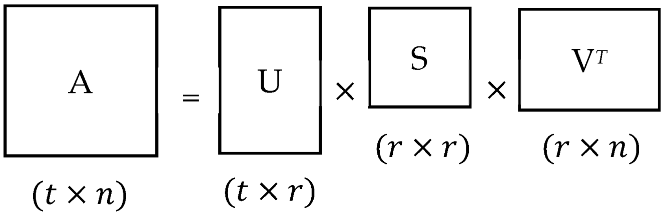

Figure 1. Singular value decomposition.

The SVD function has three matrixes: "matrix $U_{t\times r}$", "orthogonal matrix $S_{r\times r}$", and "document matrix $V^T_{r\times n}$", where t refers to the number of words, n is the document term, and r is the number of concepts in the semantic space [19].

Step 3.2: Dimension Reduction

After running SVD, it is necessary to reduce the dimensional space as it may still contain a lot of unimportant information. So as to not to affect the original characteristics, it is important to choose the feature value k. In this study, the Scree Test was used to determine the k value, as shown in Figure 2. The point before leveling was the decision point (k = 5), because the variation gap becomes smaller after the flat slope, and can be ignored after the decision point.

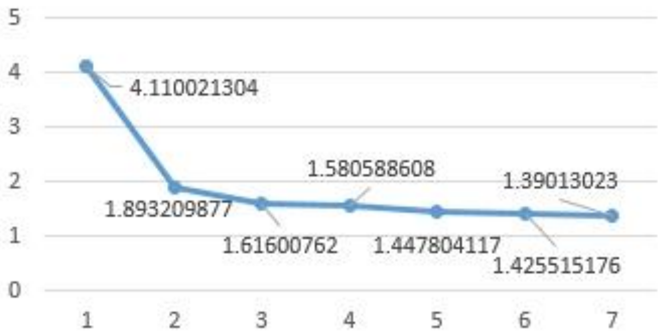

Figure 2. Scree plot.

Figure 3 shows that from the fifth value (k = 5), the eigenvalue does not vary much. Thus, we took the first five columns of the S matrix to reduce the dimensionality.

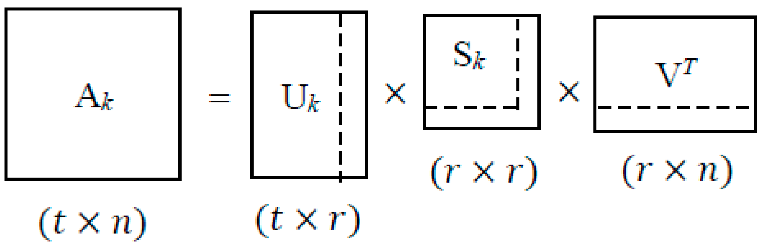

Figure 3. SVD after dimension reduction.

Step 3.3: Orthogonal rotation of axes

The concept load L_T was then calculated by multiplying the dimensionally delimited concept U_k with the concept S_k, which was calculated with Equation (2) to obtain the concept load L_T, then each feature word was ranked according to the load and the word concept was named.

$$L_T = U_k \times S_k \tag{2}$$

Step 4: Concept naming

We selected the top 30 terms in each concept. Then, based on selected important terms, we named the concepts.

Step 5: Conclusion and Discussion

Finally, based on the result of the concept naming, we presented suggestions for the reference of players and advertisers.

4. Experimental Results

4.1. Employed Data and Text Data Processing

In this study, online reviews on the Steam website were used as analysis data. We only considered the "Top Rated" games in the "EARLY ACCESS TITLES", as shown in Figure 4, because vendors were more likely to advertise unreleased games than already released games.

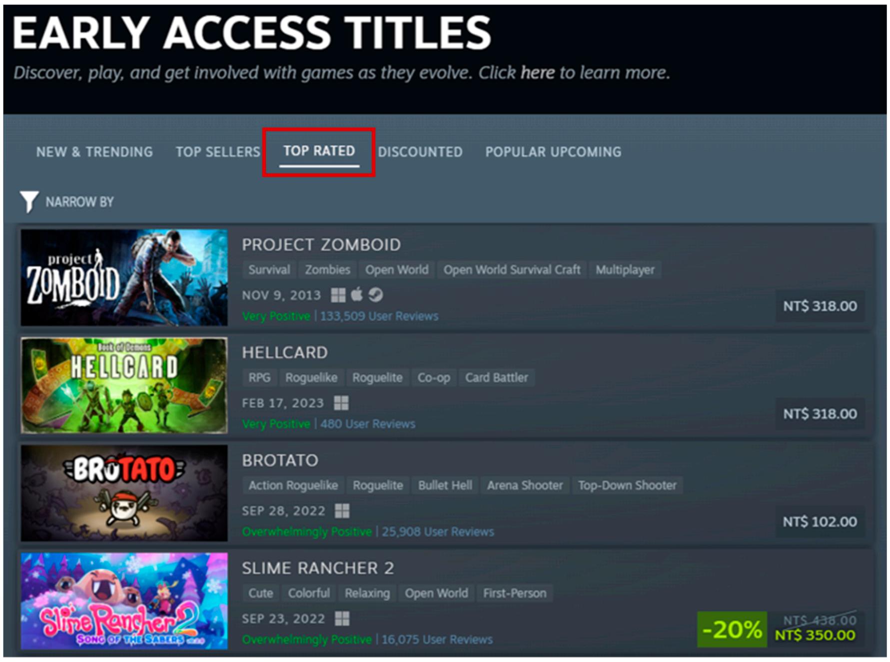

Figure 4. Highest rated game in the sneak preview version.

In addition, we selected games with "overwhelmingly positive reviews" and "recently released" among the preemptive games as shown in Figure 5. Previous studies found that the manufacturers who purchased reviews were usually unknown brands and the products were mostly young products [8].

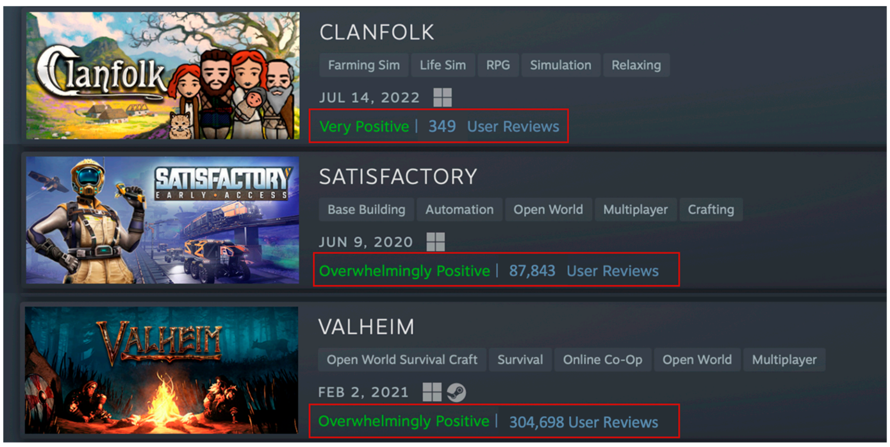

Figure 5. Collection of three games.

We selected three strategy-based games, "clanfolk", "satisfactory", and "valheim" (Figure 5) with 349, 87,843, and 304,689 reviews, respectively. Longer comments (>100 words) were more likely to be correctly classified by machines than shorter comments (<100 words) [22]. We eliminated comments with a total text word count of less than 300 and selected a final dataset of 136 comments. After processing, a 137 $\times$ 572 TDM was built.

4.2. LSA and Concept Naming

Based on the results of the LSA experiments, five concepts were determined. Keywords and their loadings in each concept are shown in Table 1 After ranking the characteristic words according to their high loadings, the top 30 words of each loading were selected to name the concept. Table 2 summarizes the selected five concepts in text reviews.

Table 1. Extracted Concepts of Keywords and Their Loadings.

Concept 1		**Concept 2**		**Concept 3**	
Keywords	Loadings	Keywords	Loadings	Keywords	Loadings
game	1.7437	rimworld	0.6508	factory	0.5840
like	0.7756	clanfolk	0.6214	factorio	0.3856
get	0.6231	game	0.4205	machine	0.3637
rimworld	0.6086	factory	0.3614	build	0.3045
thing	0.5534	friend	0.3395	velheim	0.2922
⋮	⋮	⋮	⋮	⋮	⋮
Concept 4		**Concept 5**			
Keywords	Loadings	Keywords	Loadings		
rimworld	0.3471	like	0.3795		
survival	0.3348	get	0.3022		
need	0.3191	factory	0.2704		
clanfolk	0.3092	clanfolk	0.2495		
love	0.2965	bad	0.2405		
day	0.2885	feel	0.2249		
⋮	⋮	⋮	⋮		

Table 2. Constructed Concepts.

Concept	Concept Naming	Representative Terms
1	Personal feelings	like, get, really, much, love, need, fun, feel, great, good
2	Game Content Settings	game, friend, family, winter, world, medieval, colony, multiplayer, Scottish, food
3	System Construction	factory, machine, build, server, bad, resource, production
4	Game Props	priority, food, alien, iron, tree, focus, plant
5	Playtime	bad, time, end, spend, started, since, base, playing, access, early, way, day

5. Conclusions

Using online reviews from game social media, the degree of relevance between words and meanings was determined through text mining and LSA to conceptualize the names of words and determine five elements of commercial product reviews. First of all, "personal feelings" were mentioned to make online players mistakenly think that it was a real player's comment and believe them as the commercial commentary of advertisements. In addition, the concepts of "game content settings", "system construction", and "game props" were recommended as game details to give players professional guidance. Game recommendations attracted professional players for reference. The last concept was "playtime" to recommend players to play games freely and without time restrictions. According to the results of this study, the connotations of reviews as a positive concept were mostly related to “recommendation and promotion”. If a review showed a negative concept, the review contained more information about the player’s “real thoughts”.

Author Contributions: Conceptualization, Y.-Y.W. and L.-S.C.; methodology, V.N. and Y.-Y.W.; software, V.N. and Y.-Y.W.; validation, Y.-Y.W.; formal analysis, Y.-Y.W. and L.-S.C.; writing—original draft preparation, Y.-Y.W.; writing—review and editing, V.N. and L.-S.C.; visualization, Y.-Y.W.; supervision, L.-S.C.; project administration, L.-S.C.; funding acquisition, L.-S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Science and Technology Council, Taiwan (Grant No. MOST 111-2410-H-324-006).

Informed Consent Statement: Not applicable.

Data Availability Statement: The secondary data that were used for analysis in this study are available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shehu, E.; Nabout, N.; Clement, M. The risk of programmatic advertising: Effects of website quality on advertising effectiveness. *Int. J. Res. Mark.* **2021**, *38*, 663–677. [CrossRef]
2. Maseeh, H.A.; Jebarajakirthy, C.; Pentecost, R.; Ashaduzzaman, M.D. A meta-analytic review of mobile advertising research. *J. Bus. Res.* **2021**, *136*, 33–51. [CrossRef]
3. eMarketer. Digital Advertising in 2022: Market Trends & Predictions. 2022. Available online: https://www.insiderintelligence.com/insights/digital-advertising-market-trends-predictions/ (accessed on 20 April 2022).
4. Goodwin, I. Programmatic alcohol advertising, social media and public health: Algorithms, automated challenges to regulation, and the failure of public oversight. *Int. J. Drug Policy* **2022**, *169*, 103826. [CrossRef] [PubMed]
5. Gao, C.; Zeng, J.; Lo, D.; Xia, X.; King, I.; Lyu, M.R. Understanding in-app advertising issues based on large scale app review analysis. *Inf. Softw. Technol.* **2022**, *142*, 106741. [CrossRef]
6. Mishra, A.; Satish, S.M. eWOM: Extant Research Review and Future Research Avenues. *Vikalpa* **2016**, *41*, 222–233. [CrossRef]
7. Martens, D.; Maalej, W. Towards understanding and detecting fake reviews in app stores. *Empir. Softw. Eng.* **2019**, *24*, 3316–3355. [CrossRef]
8. He, S.; Hollenbeck, B.; Proserpio, D. The Market for Fake Reviews. *Mark. Sci.* **2022**, *41*, 896–921. [CrossRef]
9. Chang, J.; Chen, M.; Chen, L.; Tseng, S. Why Customers Don’t Revisit in Tourism and Hospitality Industry? *IEEE Access* **2019**, *7*, 146588–146606. [CrossRef]
10. Samuel, A.; White, G.R.; Thomas, R.; Jones, P. Programmatic advertising: An exegesis of consumer concerns. *Comput. Hum. Behav.* **2021**, *116*, 106657. [CrossRef]
11. Meta. Fourth Quarter and Full Year 2021 Results. 2022. Available online: https://investor.fb.com/investor-news/press-release-details/2022/Meta-Reports-Fourth-Quarter-and-Full-Year-2021-Results/default.aspx (accessed on 2 February 2022).
12. Sensor Tower. Q1 2022 Data Digest: Worldwide App Downloads Reached 37 Billion Last Quarter. 2022. Available online: https://sensortower.com/blog/q1-2022-data-digest (accessed on 1 April 2022).
13. Liu, H.; Liu, S. Considering In-App Advertising Mode, Platform-App Channel Coordination by a Sustainable Cooperative Advertising Mechanism. *Sustainability* **2020**, *12*, 145. [CrossRef]
14. Dinesh, L.; Kaviya, G.; Kavya, N.C.; Kousalya, A.; Lavanya, N. Product Reviews on Opinion Mining using NLP Techniques. In Proceedings of the 2022 7th International Conference on Communication and Electronics Systems, Coimbatore, India, 22–24 June 2022; pp. 952–957.
15. Alzate, M.; Arce-Urriza, M.; Cebollada, J. Mining the text of online consumer reviews to analyze brand image and brand positioning. *J. Retail. Consum. Serv.* **2022**, *67*, 102989. [CrossRef]
16. Biswas, B.; Sengupta, P.; Kumar, A.; Delen, D.; Gupta, S. A critical assessment of consumer reviews: A hybrid NLP-based methodology. *Decis. Support Syst.* **2022**, *159*, 113799. [CrossRef]
17. Liu, Q.; Lu, J.; Zhang, G.; Shen, T.; Zhang, Z.; Huang, H. Domain-specific meta-embedding with latent semantic structures. *Inf. Sci.* **2021**, *555*, 410–423. [CrossRef]
18. Hutchison, P.D.; Daigle, R.J.; George, B. Application of latent semantic analysis in AIS academic research. *Int. J. Account. Inf. Syst.* **2018**, *31*, 83–96. [CrossRef]
19. Chen, W.; Chen, L.; Pa, Y. A text mining-based framework to discover the important factors in text reviews for predicting the views of live streaming. *Appl. Soft Comput.* **2021**, *111*, 107704. [CrossRef]

20. Das, P.; Sultana, N. Sentiment Analysis on Comments in Bengali Language Using Text Mining & Machine Learning Approach. In Proceedings of the IEEE 7th International conference for Convergence in Technology, Pune, India, 7–9 April 2022; pp. 1–4.
21. Hsiao, Y.; Hsiao, Y. Online review analytics for hotel quality at macro and micro levels. *Ind. Manag. Data Syst.* **2020**, *121*, 268–289. [CrossRef]
22. Salminen, J.; Kandpal, C.; Kamel, A.M.; Jung, S.; Jansen, B.J. Creating and detecting fake reviews of online products. *J. Retail. Consum. Serv.* **2022**, *64*, 102771. [CrossRef]