ORIGINAL ARTICLE

Let's play on Facebook: using sentiment analysis and social media metrics to measure the success of YouTube gamers' post types

Flora Poecze[1] · Claus Ebster[2] · Christine Strauss[2]

Abstract

This paper discusses the analysis results of successful self-marketing techniques on Facebook pages in the cases of three YouTube gamers: PewDiePie, Markiplier, and Kwebbelkop. The research focus was to identify significant differences in terms of the gamers' user-generated Facebook metrics and commentary sentiments. Analysis of variance (ANOVA) and k-nearest neighbor sentiment analysis were employed as core research methods. ANOVA of the classified post categories revealed that photos tended to show significantly more user-generated interactions than other post types, while, on the other hand, re-posted YouTube videos gained significantly fewer numbers in the retrieved metrics than other content types. K-nearest neighbor sentiment analysis pointed out underlying follower negativity in cases where user-generated activity was relatively low, thereby improving the understanding of the opinion of the masses previously hidden behind metrics such as the number of likes, comments, and shares. The paper at hand highlights the methodological design of the study as well as a detailed discussion of key findings and their implications, and future work. The results per se indicate the need to utilize natural language processing techniques to optimize brand communication on social media and highlight the importance of considering machine learning sentiment analysis techniques for a better understanding of consumer feedback.

Keywords Social media metrics · Self-marketing · Sentiment analysis · Online gaming

1 Introduction

Self-marketing as an activity to promote a person (as opposed to a product or service) was recognized as early as the 1960s by marketing scholars such as Kotler and Levy [1], but the emergence of Web 2.0 and the birth of social networks shortly after the millennium opened a unique opportunity for individuals' self-representation and self-expression [2]. YouTube became the market leader in video content sharing, and it did so by providing content creators with tools and self-marketing techniques [3] that allowed them to successfully gain

✉ Christine Strauss
christine.strauss@univie.ac.at

Flora Poecze
1719001127@fh-burgenland.at

Claus Ebster
claus.ebster@univie.ac.at

[1] University of Applied Sciences Burgenland, Campus 1, 7000 Eisenstadt, Austria

[2] University of Vienna, Oskar Morgenstern Platz 1, 1090 Vienna, Austria

monetary profits from other people viewing their content as well as possible worldwide recognition [4–6].

The birth of YouTube gaming and the increasing popularity of gaming channel owners have encouraged scientific researchers to conduct in-depth studies on the characteristics of self-marketing communication and its effectiveness in the content forms that grab user attention and those that lead to a lack of follower activity. Thus, the present quantitative exploratory study introduces an approach to measure the communication effectiveness of YouTube gamers by analyzing the content posted on their Facebook brand pages.

The goal of this paper (which is an extended version of [7] presented at the 9th International Conference on Ambient Systems, Networks and Technologies) is to identify the relationship of these content types with user-generated metrics on Facebook, such as the number of likes, comments, and shares, complementing these results with sentiment analysis of the commentaries' content appearing below the sampled posts.

The results provide valuable findings that can be used by companies and self-marketed individuals to improve their brand communication on social media. Furthermore, by using this approach, independent game developers and actors in the ever-growing gaming industry gain valuable insights from

their customers' feedback to enhance their future products as well as their promotion.

2 Background and related work

As social media and the associated metrics fall into the intersecting area of management and informatics, this section provides related work in the context of (self-)marketing and related work in the context of social media and sentiment analysis.

2.1 Self-marketing: background and related work

Research on self-marketing as a distinct type of marketing communication that applies classic branding techniques to human beings [2, 6] began in the 1960s [1] and was revisited in the late 1990s [8]. However, a gap exists between researchers and practitioners (consultancy services, self-improvement groups) in this domain, the latter dominating the area [6]; therefore, researchers have increasingly advocated the need for scientific studies to further examine the field of self-marketing [3, 6, 9].

The terminology of this phenomenon is the subject of ongoing discussion. Terms such as "personal branding," "personal marketing," and "self-marketing" have been used interchangeably [3]. Individuals as actors of this phenomenon are referred to as "human brands," "branded individuals," and "branded personas" [3, 10]. The extant literature in this domain focuses on the process of personal branding [3, 5, 6, 11] and aims to determine the underlying factors of success by branded persons.

Studies discussing the importance and role of sociocultural background in fruitful personal branding [3, 6, 11] have considered Bourdieu's field theory [12] and theses about the forms of capital [13]. YouTube functions as an organized field where YouTube gamers are actors (referred to as agents by Bourdieu [12], each of whom holds cultural capital in the form of educational qualifications and social capital as defined by their acquaintances and social networks). Each channel owner has a unique habitus [12] as well, which consists of attitudes and behavioral characteristics that are crucial to the owner's self-representation, for example, in his or her YouTube videos.

The process of self-representation [14] involves the channel owner's performance being watched by observers in a front (i.e., his or her YouTube channel); it generally consists of a video background setting and the owner's manner and appearance provide great opportunities for aspiring gamers [2] to "stand out" [6] and offer a unique experience for the audience.

YouTube gamers have been gaining in popularity [2, 6] in the last decade, and studies have pointed out numerous benefits of social media data analysis [15], indicating that the analysis of user-generated text commentary, or the online opinion-mining of the masses, has become one of the most pressing

issues [16, 17]. User-generated content (UGC) as a means for user communication and self-expression has been studied in the social media context for more than a decade [7, 18, 19]. As a part of the growing body of research in the UGC domain, social network sites are garnering particular research interest, especially in the area of brand promotion [20].

2.2 Sentiment analysis and social media: background and related work

The emergence of an extensive amount of digitalized and online publicly available opinion data has encouraged researchers to find solutions for efficient analysis, which opened the research area of sentiment analysis. While companies (e.g., Microsoft, Google, SAS, Hewlett-Packard) often build their in-house capabilities [17], application-oriented research papers have emerged in various fields as well, election outcomes [21–23], movie box office revenues [24, 25], and stock prediction [26, 27] being among the most widely studied domains. However, due to the relative infancy of this research domain, determination of the most suitable methodology is still under discussion.

Sentiment analysis, often called opinion mining is a research area that focuses on analyzing written texts about sentiments, appraisals, attitudes, and opinions of people according to certain entities (products, organizations, services, issues, events, topics, or other individuals) [28, 29].

Although sentiment analysis is a very active research field of natural language processing (NLP), its roots go back to the 1980s. Its forerunners are considered to be projects investigating "belief analysis." Both Wilks [30] and Carbonell [31] generated a computer system for analyzing subjective understanding. During the late 1980s and 1990s, studies were mainly focusing on the interpretation of narrative [32, 33] directionality of a sentence [34], and point of view (POV) tracking: the storyteller's POV [35] or a certain character's psychological POV [36, 37]. Hatzivassiloglou and McKeow studied adjectives joined with conjunctions and explored the polarity of the adjectives and the effect of conjunctions on the sentiment of these adjectives [38].

The research field primarily focuses on written text; due to this, it is mainly studied in the field of NLP. Nevertheless, it is also a subject of web mining, information retrieval, and data mining of text data [28].

In the last two decades, social media have become a defining feature of the Internet. The characteristics of online social media can be grasped according to three constitutive elements first described by Ellison [39]. First, there has to be a main platform, a virtual space, where users can create their profiles, thereby representing themselves to other users of the environment. The second crucial element is network creation which helps users communicate with each other. Lastly, public visibility of a user's personal network (i.e., friends) is required to enable users to wander to other users' network graphs.

Since the advent of social media, digital, opinionated data has emerged on the Internet, which reshaped information retrieval and had a significant impact on individuals, companies, and governments. Although the Internet contains an immense volume of opinionated texts (social platforms, review sites, blogs, forums), all of the above mentioned entities faced the problem of how to obtain necessary information from them, which led to the development of sentiment analysis that can be divided into "real-world" applications and application-oriented research papers [28].

A turning point in the application of sentiment analysis to social media was a study by Asur and Huberman [40] analyzing Twitter public sentiment, after which an increasing number of studies using various social media platforms to acquire the sentiment of the crowds emerged (cf. [41]). The common goal of this and similar studies in this research domain is that they aim to build models according to past opinionated data and try to refine this model in such manner that it can be used for predictive purposes too.

These studies highlighted a crucial element to the authors of this paper, namely the importance of in-depth analysis of metric data. For instance, in a study predicting movie sales based on the sentiment expressed toward movies on social media, Mishne and Glance [42] used metrics such as income per screen or raw sales; Liu et al. [43] raised awareness on the importance of time-stamps; Asur and Huberman [40] also highlighted the relevance of such meta-data.

Another application of sentiment analysis on social media is political marketing and the prediction of elections. The main reason why the topic of election prediction and its emergence since 2010 has become a popular research area is that researchers had begun investigating cheaper possibilities instead of the traditional, costly polling techniques [44, 45]. The first studies appearing in this showed promising results area [44, 46]. Tumasjan et al. [46] investigated the German Elections of 2009 and tried to predict its results with aggregated tweets that mentioned one of the six major German parties. Their results strongly correlated with the results of the election, which urged researchers to try to replicate their methodology [21–23].

In [21] the main difference between election prediction from public sentiment and traditional polling techniques is pointed out: the latter has existed for 80 years and is capable of avoiding sampling bias, while weighting a sample of Twitter data according to the gender, age, or other demographic factors of tweeters is not possible. The very same argument also appeared in [47].

Although there are numerous different metrics that can be scraped from social networks, there is a difference between quality metrics and the so-called "vanity metrics," which may reflect one's successful content marketing, but are rather misguiding, mentioning Facebook likes, Twitter follower counts, or page views as examples. On the other hand, quality metrics deliver useful information on a deeper level: sentiment analysis is considered one of them, providing data about the attitudes and emotions of users through the analysis of their written communication on brand pages [48].

Research according to sentiment analysis in social media can be divided into two research streams: while most of the studies concentrated on natural language processing of texts in social media, their methodology is limited only to the textual information [49, 50]. Other approaches are emerging in recent literature that attempt to overcome these limitations by trying to extract information such as friendships and connections among users, pointing out, that connected users are likely to have similar opinions regarding a particular subject [17, 51].

We support the recently emerging studies that have argued for the complementary use of sentiment analysis of social media [17] and its "traditional" retrievable metrics (i.e., number of likes, comments, and shares of posted content) to achieve a deeper understanding of audience reactions to communication forms of self-marketing on social media.

Thus, this study has three primary research questions (RQs): exploration of the relationship of sampled posts' content types and their user-generated metrics (RQ1), the relationship of sampled posts' content types and their user-generated commentary sentiment results (RQ2), and the comparison of these two analyses (RQ3) by exploring the relation of these user-generated metrics and the sentiment results.

By the evaluation and discussion of these research questions, we are in line with the call for further research in this domain [3, 6, 9]. We intend to provide further insights into this area of interest, which is rapidly developing and therefore needs appropriate and well-founded methods and tools.

3 Method

The following subsections reveal in detail the activities performed during the various phases of the study, i.e., sampling, data collection, and data analysis. The chosen criteria, tools, and methods, which we applied in various steps of the analysis, are shown as a synopsis of the study's methodological layout in Table 1.

3.1 Sampling

As units of analysis in the first stage of sampling, which can be considered a judgment sample, three YouTube gamers were chosen: PewDiePie, Markiplier, and Kwebbelkop. These gamers share common characteristics by often playing the same games, being in the top 100 most popular gaming channels, and frequently using Facebook as a means of communication. Furthermore, they were part of Revelmode, a sub-network of Maker Studios owned by Disney.

Table 1 Overview on the study's methodological layout: applied criteria, tools, and methods in various phases of study

Phase	Activity	Criteria/tool/method
Sampling	Determine 3 gamers for analysis	•Among top 100 gamers worldwide •Offering "Let us play"-games •Using Facebook as primary channel
	Choose 50 posts from each gamer	Time
Data collection	Retrieve all comments associated with chosen posts	Applied tool: Netvizz (i.e., integrated Facebook application)
Data analysis	Classify posts	Grounded theory
	Analyze Facebook metrics	•Univariate data analysis •Kruskal-Wallis H test •Dunn's post-hoc
	Conduct sentiment analysis of comments	•K-nearest neighbor •Supervised learning approach •Kruskal-Wallis H test •Dunn's post-hoc
	Compare "content types" versus "Facebook metrics"	Descriptive comparison

Criterion sampling was used during the second stage of sampling [52]. The same number of posts ($n = 50$) with the same end point on the sampling time scale was established to allow for comparisons among the sampled YouTubers.

Last, in the third sampling stage, all comments appearing under the sampled posts were retrieved for future sentiment analysis purposes.

3.2 Data collection

To collect data for our study we used a tool, an integrated Facebook application, i.e., Netvizz (version 1.45) [53], which is capable of scraping data about groups, pages, page like networks, page timeline images, and search and link statistics. Netvizz was used to extract 50 posts from each of the pages of the sampled YouTubers and their Facebook metrics, along with all comments on those posts with timestamps. Netvizz uses the public application program interface (API) of Facebook for data retrieval and can retrieve solely those data types that are available through the API [53].

3.3 Classification of posts

Grounded theory was used to classify the retrieved 150 posts [54]. Following the coding mechanism of this theory, open coding consisted of categorization of the retrieved posts into four core categories: link, photo, status update, and video.

During axial coding, the previously generated core categories were revised. As a result, the core category "video" was divided into two subcategories: "integrated Facebook videos" that were made explicitly for the Facebook audience and "embedded YouTube videos" that first appeared on YouTube and were later reposted on Facebook.

After successful determination of core categories, the sampled posts were further classified into subcategories according to their content during the selective coding phase. Following grounded theory, these subcategories were subservient to the previously generated core categories [54].

3.4 Data analysis methods

After univariate analysis of the retrieved user-generated metrics, analysis of variance (ANOVA) was performed to test for possible significant differences among the classified content types in terms of their Facebook metrics. As a result of Kolmogorov-Smirnov tests to decide between non-parametric and parametric analysis approaches, Kruskal-Wallis H tests were conducted along with Dunn's post-hoc pairwise comparisons [55].

The supervised learning method k-nearest neighbor (k-nn) was chosen for the sentiment classification of the retrieved commentaries. Sentiment classification can be achieved with supervised (e.g., naïve Bayes, support vector machine, k-nn) and unsupervised (sentiment dictionaries, such as, e.g., SentiWordNet) learning approaches [17]. We considered that the international gaming community exhibits a specific "lingo" [56] which makes it rather difficult to analyze with sentiment dictionaries. For example, the slang words "noob" (newcomer), "frag" (to virtually kill another player), "flaming" (to verbally attack), and "rekt" (wrecked, destroyed) are widely used in this community but are challenging to accurately classify by a sentiment dictionary. Therefore, we employed a supervised learning approach and used a training set with hand-labeled commentaries as its items.

Due to the rarity of neutrality regarding comment sentiments on social media [57, 58], a bivariate training set consisting of an equal number of hand-labeled positive and negative commentaries was used. This set was "trained" on

the test data (i.e., the remaining comments), where machine-based classification was performed according to the rules that the computer "learned" from the training set [16, 59].

The aforementioned *k*-nn approach predicted the sentiment score of a test data item according to its similarities to previously tagged training set items. The sentiment score was determined in the same manner as the training set item that was most comparable to it, using the most similar word patterns [59]. The analysis was performed with varying amounts of nearest neighbors (i.e., closest matching items of the training set) to achieve the most accurate prediction of the data set items [59]. For this reason, we also tested accuracy with alternative numbers of positive and negative training set items ($N = 100$; $N = 200$; $N = 400$). The process was conducted using the educational license of RapidMiner Studio 7.5 [60], which is a code-free environment for designing advanced analytic processes with machine learning, data mining, text mining, predictive analytics, and business analytics.

Similar to the Facebook metrics analysis, Kruskal-Wallis H tests with Dunn's post-hoc pairwise comparisons were used to determine possible significant differences among the sentiment score means of the distinct post categories.

The results of the Facebook metrics and sentiment analysis were compared in the last step of the data evaluation to reveal similarities and differences.

4 Results

Our results will be presented in three separate sections. First, the Facebook metrics analysis will reveal distinctive differences in user engagement regarding both core and subcategories of the retrieved posts by all analyzed YouTube gamers, followed by a successful ANOVA regarding their commentary sentiments as well. Last, the results of both sections will be compared, highlighting the possible value and necessity of conducting both analyses in practice.

4.1 Facebook metrics analysis

Classified core categories of the sampled posts displayed distinctive differences. While Kwebbelkop almost exclusively posted photos, in the samples of PewDiePie and Markiplier reposts of previously published YouTube videos were in the majority.

Significant differences in core post categories and their Facebook metrics were detected between like, reaction, and total engagement scores of all three YouTubers analyzed. However, regarding shares, only Markiplier's sample showed significant differences in the core categories of the sampled posts. Differences in the means of Kwebbelkop's comment scores were also non-significant. Figure 1 summarizes the rank orders of the classified core categories by the sampled

YouTubers where the abovementioned significant differences were observable. (The figure uses numeric differentiation for representation of the classified post types: 1, link; 2, photo; 3, status update; 4, integrated Facebook video; 5, embedded YouTube video).

The ANOVA results of the core post categories and their Facebook metrics revealed that links tended to receive significantly fewer user interactions than did photos. The same patterns were observed in the case of embedded YouTube videos. This core category also received significantly less user interaction by all measured Facebook metrics than did photos. These results are especially interesting in the light of previously discussed ratios of sampled posts in different core categories. In the cases of Markiplier and PewDiePie, the number of reposted YouTube videos was larger than any other core post categories. However, their shared photos received significantly more user-generated interactions. These highly engaging photo posts accounted for 34% of posts sampled from Markiplier's page, whereas PewDiePie's sample contained only two photos, despite their excessive popularity among his followers.

The ANOVA of post subcategories and their Facebook metric means demonstrated that photos depicting family, friends, pets, and/or the YouTubers themselves received significantly more user-generated actions than subcategories with the lowest means in the sample. Furthermore, posts encouraging audience interaction and engagement (e.g., give an opinion, "like" if agree) received significantly higher Facebook metric means than lower ranking subcategories.

As Fig. 2 illustrates, the rank orders in terms of their share and comment means changed distinctively. The figure uses the same metric system for the subcategory visualization as in Fig. 1 (the first number represents the core post category, whereas the second number refers to the subcategory of this particular core category). As an example, in the case of Markiplier, some posts discussed the problems regarding YouTube and the gamer community (i.e., the subcategory numbered "15"). As seen from Fig. 2, this subcategory reached outstanding results regarding its "share" metrics and also ranked higher in the number of comments it gained than in "likes" or even "total engagement" scores. The *k*-nn sentiment analysis of user commentaries provides valuable insights into the reasons why these posts were widely shared and commented, although they did not gain an extensive number of likes.

4.2 Sentiment analysis

The accuracy test of the performed *k*-nn sentiment analysis, conducted with RapidMiner Studio 7.5, resulted in 82.3% accuracy using a training set containing 200 positive and 200 negative comments, with $k = 2$ nearest neighbors.

After the successful sentiment classification, Kruskal-Wallis H tests were conducted. Significant differences in the

Fig. 1 Core category median rankings of the analyzed Facebook metrics by the sampled YouTubers

mean sentiment scores of both core and subcategories of the retrieved posts were detected for all three YouTubers.

Embedded YouTube videos tended to gain significantly less positivity in their sentiment means throughout the sample than any other core categories. In comparison to other content types, the ratio of negative commentaries submitted for this post category was relatively high.

In contrast to re-posted YouTube videos, photos tended to receive significantly more positive sentiments than other core categories for all three gamers, with the highest ratio of positive commentaries in the samples.

The subcategory sentiment analysis revealed that although likes for certain categories had relatively low means, they "jumped" to the top-ranking places in terms of positivity means. The aforementioned discussion of commentary regarding YouTube problems and other events that shook the gaming community during the sampling time had a high effect on user commentary sentiments, where fans tended to lend their support regarding these discussed issues.

4.3 Comparison of Facebook metrics and sentiment analyses

The post category results comparison regarding Facebook metrics and sentiment means highlighted the relevance of commentary sentiment observation and its complementary function in understanding content popularity.

With respect to photos, although sentiment positivity accentuated the preliminary Facebook metrics results and showed that posts in this core category do not merely receive a large number of user-generated actions, they also stimulated positive written opinions.

Fig. 2 Subcategory median rankings of the analyzed Facebook metrics by the sampled YouTubers

However, in particular cases, sentiment analysis revealed hidden feedback negativity and/or possible debates in the comment sections, manifested by the metrics and sentiment comparison of *self-sponsoring* posts, analyzed in PewDiePie's sample. During the sampling timeframe, PewDiePie introduced his mobile app "Legend of the Brofist," which was introduced first to the Apple and Android app store, leaving Microsoft phone users to wait two and a half months longer for its release. Therefore, self-sponsoring posts containing the app's advertising received a relatively high ratio of negative comments, which was not previously detectable by the mere analysis of its user-generated Facebook metrics.

The application of user commentary sentiment analysis emphasized the Facebook metrics results; as discussed in the case of photos, it proved to be a useful complementary tool to expose counter-opinions, originally hidden by the mere analysis of user-generated metrics.

5 Conclusion and discussion

The growing importance of social media as a unique tool for self-marketing was the subject of this analysis, investigated in the case of YouTube gamers leveraging Facebook with the intention of communicating with their audience and fans. We proposed three research questions to determine whether content types differ significantly in terms of their user-generated metrics (RQ1), sentiment analysis results (RQ2), and how these findings can be compared (RQ3). The results suggest the importance of not relying on and using solely retrievable user-generated metrics. Instead, the sentiment of the text commentary accompanying social media posts should also be incorporated in the analysis.

5.1 Key findings

In the cases of PewDiePie and Markiplier, the core content in the sampling time interval consisted of reposted YouTube videos. However, the relative lack of user activity paired with a high ratio of negative user commentaries called into question such a communication strategy. On the other hand, photos, especially those showing friends, family, pets, and the YouTube gamers themselves, resulted in the highest user activity and positive comment ratios.

Sentiment analysis of Facebook metrics revealed the unpopularity of certain content and underlying fan debates and disagreements with the gamers' opinions about frequently discussed issues. Such controversy during the sampling time interval was caused by the H3H3 copyright infringement case [61] where Markiplier voiced his opinion and took a stand by the content creators of H3H3 in one of his YouTube videos, reposting it to Facebook where a strong discussion began in the comment section, detected during the analysis. Another

example for debate was the case of the game called "Bear Simulator," [62] developed by first-time independent game developers after a successful Kickstarter campaign, played and posted by PewDiePie during the sampling timeframe. In his Let us Play video about it, PewDiePie made remarks about the poor quality of the game and commenters also voiced their negative opinions about it. This particular Facebook post received an extensive amount of negative commentary compared to other re-posted YouTube contents, whereas it received relatively low reactions regarding its Facebook metrics. By using the k-nn approach, we could explain the reasons behind the striking amount of comments in comparison to other metrics. Machine-learning based sentiment analysis proved to be capable of analyzing hundreds and thousands of commentaries accompanying a single post.

5.2 Managerial implications

The combined analysis of user-generated metrics and sentiment classification provides information that is necessary for self-marketed individuals to optimize their communication on social media, which plays an important role in achieving steady audience growth that can result in monetary profits. Application of these techniques allows independent game developers and the gaming industry as a whole to gain insights into the critical reception of their products from the target group itself. In consideration of the steadily growing global phenomenon of social media marketing and the growing number of companies with diverse products and services targeting potential consumers on various social media platforms, the authors of this paper believe that both user-generated metrics analysis and sentiment mining are essential for marketers in the contemporary era.

5.3 Limitations and further research

Future studies should clarify the weights and roles of the communication effectiveness indicators used in this research project by employing a larger sample and possibly stepping out of the domain of YouTube gamer brand personalities. This will help to improve the study of brand personality communication on social media, regardless of the domain in which these persons are active.

Recent studies on behavioral motivations in social media engagement have shown the importance of the uses and gratification theory [63, 64] and social exchange theory [65, 66] for the deeper understanding of factors that influence users to share, like, or comment on social media posts. In this research direction, the audience of the discussed self-marketed gamers inspected as a potential virtual community can lead to an analysis of the antecedents that motivate the sense of virtual community (SOVC) from the uses and gratifications aspect [67]. Furthermore, the presence of the discussed gamers on multiple

social media platforms (e.g., Twitch.tv, Twitter, Instagram) poses the question of the extent (or ratio) to which they are followed by the same audience members on different networks. Recent studies have found tension release as a strong positive predictor of how many hours users watched Twitch.tv online gaming streams, the number of streamers followed, and also the number of streamers watched [68]. Moreover, a novel study [69] concluded that affective motivations are evoked by the content type of the stream on Twitch.tv rather than the game genre being broadcast. On the other hand, the analysis of game genres revealed that real-time strategy (RTS) games (e.g., Age of Empires III, Stellaris) show a negative association with affective motivations, while other game genres evoke a tension release; these are multiplayer online battle arena (MOBA) games (e.g., League of Legends, Dota 2), collectible card games (CCG; e.g., Hearthstone), and first-person shooter (FPS) games (e.g., Tom Clancy's Rainbow Six Siege, Doom, Overwatch). Further studies should explore whether these results can be extended for YouTube gamers and their audience.

The investigation of user motivational factors combined with the presently described post classification method, Facebook metrics analysis, and sentiment-mining technique can help determine possible motivational differences in user engagement in the analyzed post types. Further research should clarify the possible interrelationships between underlying behavioral dimensions and the extent of user engagement in terms of different post types and examine their relevance from economic, social psychological, and marketing aspects.

This approach used the supervised learning method of sentiment analysis, k-nn. Although the test can predict the commentary sentiments with 82.3% accuracy, their possible review using other techniques (e.g., naïve Bayes, support vector machine, or sentiment dictionaries such as SentiWordNet) would add further insights to the present findings.

Our study used a bivariate training set regarding commentary sentiments by labeling positive and negative comments. However, sentiment description can be extended by sublevels as well (e.g., joy, love, and surprise being positive and anger, sadness, and fear being negative) [70]. Adding these emotional sublevels as means for sentiment description will likely extend the presented results.

The application of various methods and/or their combination supports successful (self-)marketing activities in social media environments and also provides valuable insights for gaming industry actors and social media marketing research professionals.

References

1. Kotler P, Levy SJ (1969) Broadening the concept of marketing. J Mark 33:10–15. https://doi.org/10.2307/1248482
2. Chen CP (2013) Exploring personal branding on YouTube. J Internet Commer 12:332–347. https://doi.org/10.1080/15332861.2013.859041
3. Shepherd IDH (2005) From cattle and coke to Charlie: meeting the challenge of self marketing and personal branding. J Mark Manag 21:589–606. https://doi.org/10.1362/0267257054307381
4. Pace S (2008) YouTube: an opportunity for consumer narrative analysis? Qual Mark Res 11:213–226. https://doi.org/10.1108/13522750810864459
5. Labrecque LI, Markos E, Milne GR (2011) Online personal branding: processes, challenges, and implications. J Interact Mark 25:37–50. https://doi.org/10.1016/j.intmar.2010.09.002
6. Khedher M (2012) A brand for everyone: guidelines for personal brand managing. J Glob Bus Issues 9:19–27
7. Poecze F, Ebster C, Strauss C (2018) Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts. In: Proc. 9th Int. Conf. On ambient systems, networks and technologies, ANT-2018 Procedia computer science 130:660-666. doi: https://doi.org/10.1016/j.procs.2018.04.117pp.660-666
8. Peters T (1997) The brand called you. Fast Co 10:1–8. https://doi.org/10.1093/mind/102.405.1
9. Nessmann K (2010) Personal branding and the role of public relations. Adv Advert Res 1:366–395
10. Thomson M (2006) Human brands: investigating antecedents to consumers' strong attachments to celebrities. J Mark 70:104–119. https://doi.org/10.1509/jmkg.70.3.104
11. Parmentier MA, Fischer E, Reuber AR (2013) Positioning person brands in established organizational fields. J Acad Mark Sci 41: 373–387. https://doi.org/10.1007/s11747-012-0309-2
12. Bourdieu P (1993) The field of cultural production. Columbia University Press, New York
13. Bourdieu P (1986) The forms of capital. Educ Cult Econ Soc:46–58. https://doi.org/10.1002/9780470755679.ch15
14. Goffman E (1956) The presentation of self in everyday life. University of Edinburgh, Social Sciences Research Centre, Edinburgh
15. Salampasis M, Paltoglou G, Giachanou A (2014) Using social media for continuous monitoring and mining of consumer behaviour. Int J Electron Bus 11:85–96. https://doi.org/10.1504/IJEB.2014.057905
16. Liu B (2012) Sentiment analysis and opinion mining. Synth Lect Hum Lang Technol 5:1–167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016
17. Pozzi FA, Fersini E, Messina E, Liu B (2016) Sentiment analysis in social networks, 1st edn. Morgan Kaufmann Publishers
18. Boyd DM, Ellison NB (2008) Social network sites: definition, history, and scholarship. J Comput Commun 13:210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

19. Kaplan AM, Haenlein M (2010) Users of the world, unite! The challenges and opportunities of social media. Bus Horiz 53:59–68. https://doi.org/10.1016/j.bushor.2009.09.003

20. Smith AN, Fischer E, Yongjian C (2012) How does brand-related user-generated content differ across YouTube, Facebook, and Twitter? J Interact Mark 26:102–113. https://doi.org/10.1016/j.intmar.2012.01.002

21. Gayo-Avello D, Metaxas P, Mustafaraj E (2011) Limits of electoral predictions using Twitter. In: Proc. 5th Int. AAAI Conf. Weblogs Soc. Media pp 490–493

22. Chung J, Mustafaraj E (2011) Can collective sentiment expressed on Twitter predict political elections? In: Proc. 25th AAAI Conf. Artif. Intell. p 1770–1771

23. Jungherr A (2016) Twitter use in election campaigns: a systematic literature review. J Inf Technol Polit 13:72–91. https://doi.org/10.1080/19331681.2015.1132401

24. Krauss J, Nann S, Simon D, Gloor PA, Fischbach K (2008) Predicting movie success and academy awards through sentiment and social network analysis. In: Proc. 16th Eur. Conf. Inf. Syst. pp 2026-2037

25. Joshi M, Das D, Gimpel K, Smith NA (2010) Movie reviews and revenues: an experiment in text regression. Hum Lang Technol Annu Conf North Am Chapter Assoc Comput Linguist:293–296

26. Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. J Comput Sci 2:1–8. https://doi.org/10.1016/j.jocs.2010.12.007

27. Oliveira N, Cortez P, Areal N (2017) The impact of microblogging data for stock market prediction: using Twitter to predict returns, volatility, trading volume and survey sentiment indices. Expert Syst Appl 73:125–144. https://doi.org/10.1016/j.eswa.2016.12.036

28. Liu B (2015) Sentiment analysis: mining opinions, sentiments, and emotions. Cambridge University Press

29. Nasukawa T, Yi J (2003) Sentiment analysis: capturing favorability using natural language processing. In: Proc. 2nd Int. Conf. Knowledge capture pp 70–77

30. Wilks Y, Bien J (1983) Beliefs, points of view, and multiple environments. Cogn Sci 7:95–119. https://doi.org/10.1016/S0364-0213(83)80007-X

31. Carbonell J G (1979) Subjective understanding: computer models of belief systems (no. RR-150). Yale Univ New Haven Conn Dept of computer science

32. Wiebe J M, Rapaport W J (1988) A computational theory of perspective and reference in narrative. In: Proc. 26th annual meeting on Association for Comp Linguistics pp. 131-138

33. Wiebe J M (1990) Identifying subjective characters in narrative. In: Proc. 13th Conf. on Comp. linguistics 2:401-406

34. Hearst M (1992) Direction-based text interpretation as an information access refinement. Text-based intelligent systems: current research and practice in information extraction and retrieval, pp. 257-274

35. Sack W (1994) On the computation of point of view. In: AAAI Proc. 12th Nat. Conf. Artificial Intelligence, p. 1488

36. Wiebe JM (1994) Tracking point of view in narrative. Computational Linguistics 20(2):233–287

37. Wiebe JM, Bruce RF (2001) Probabilistic classifiers for tracking point of view. Progress in Comm Sci:125–142

38. Hatzivassiloglou V, McKeown K R (1997) Predicting the semantic orientation of adjectives. In: Proc. 8th Conf. Europ. Chapter of the Association for Computational Linguistics. pp. 174–181

39. Ellison NB (2007) Social network sites: definition, history, and scholarship. J Computer-Mediated Comm 13:210–230

40. Asur S, Huberman B A (2010) Predicting the future with social media. In: Int Conf Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM Vol. 1, pp. 492–499

41. Rui H, Liu Y, Whinston A (2013) Whose and what chatter matters? The effect of tweets on movie sales. Decis Support Syst 55:863–870. https://doi.org/10.1016/j.dss.2012.12.022

42. Mishne G, Glance N S (2006) Predicting movie sales from blogger sentiment. In: AAAI spring symposium: computational approaches to analyzing weblogs. Pp. 155-158

43. Liu Y, Huang X, An A, Yu X (2007) ARSA: a sentiment-aware model for predicting sales performance using blogs. In: Proc. 30th annual international ACM SIGIR Conf. Res. and Development in information retrieval pp. 607-614

44. O'Connor B, Balasubramanyan R, Routledge B R, Smith N A (2010) From tweets to polls: linking text sentiment to public opinion time series. In: Proc. 4th Conf. on Weblogs and Social Media (ICWSM) 11:122-129

45. Bermingham A, Smeaton A F (2011) On using Twitter to monitor political sentiment and predict election results. In: Proc workshop on sentiment analysis where AI meets psychology (SAAIP) p. 2-10

46. Tumasjan A, Sprenger T O, Sandner P G, Welpe I M (2010). Predicting elections with Twitter: what 140 characters reveal about political sentiment. In: Proc. 4th intern. AAAI Conf. Weblogs and social media (ICWSM) 178-185

47. Sang E T K, Bos J (2012) Predicting the 2011 dutch senate election results with twitter. In: Proc. Workshop on Semantic Analysis in Social Media. pp. 53–60

48. Pang B, Lee L (2008) Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2(1–2):1–135

49. Maas A L, Daly R E, Pham P T, Huang D, Ng A Y, Potts C. (2011) Learning word vectors for sentiment analysis. In: Proc 49th annual meeting of the Association for Computational Linguistics: human language technologies. Pp. 142-150

50. Wang S, Manning C D (2012) Baselines and bigrams: simple, good sentiment and topic classification. In: Proc 50th annual meeting of the Association for Computational Linguistics pp. 90-94

51. Wilson T, Wiebe J, Hoffmann P (2005) Recognizing contextual polarity in phrase-level sentiment analysis. In: Proc. Conf. on human language technology and empirical methods in natural language processing. pp. 347–354

52. Bryman A (2015) Social research methods, 5th edn. Oxford University Press, Oxford

53. Rieder B (2013) Studying Facebook via data extraction. In: Proc. 5th Annu. ACM Web Sci. Conf. - WebSci '13 pp 346–355

54. Strauss AL (1987) Qualitative analysis for social scientists. doi: https://doi.org/10.1017/CBO9780511557842

55. Dunn OJ (1964) Multiple comparisons using rank sums. Technometrics 6:241–252. https://doi.org/10.1080/00401706.1964.10490181

56. Barbieri F, Espinosa-Anke L, Ballesteros M, Soler-Company J, Saggion H (2017) Towards the understanding of gaming audiences by modeling twitch emotes. In: Proc. 3rd work. Noisy user-generated text pp 11–20

57. Godbole N, Srinivasaiah M, Skiena S (2007) Large-scale sentiment analysis for news and blogs. In: Proc. Conf. Weblogs Soc. Media (ICWSM 2007) pp 219–222

58. Troussas C, Virvou M, Espinosa KJ, Llaguno K, Caro J (2013) Sentiment analysis of Facebook statuses using naive Bayes classifier for language learning. In: Proc. 4th Int. Conf. Information, Intell. Syst. Appl. pp 198–205

59. Kotu V, Deshpande B (2014) Predictive analytics and data mining: concepts and practice with RapidMiner. Morgan Kaufmann Publishers

60. Mierswa I, Wurst M, Klinkenberg R, Scholz M, Euler T (2006) YALE: rapid prototyping for complex data mining tasks. In: Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. pp 935–940

61. Vogele J (2017) Where's the fair use? The takedown of Let's play and reaction videos on YouTube and the need for comprehensive DMCA reform. Touro Law Rev 33:589–631

62. Fahey M (2016) Successful $100,000 Kickstarter dev calls it quits due to drama. From: kotaku.com/successful-100-000-kickstarter-dev-calls-it-quits-due-1763162848

63. Baek K, Holton A, Harp D, Yaschur C (2011) The links that bind: uncovering novel motivations for linking on Facebook. Comput Hum Behav 27:2243–2248. https://doi.org/10.1016/j.chb.2011.07.003

64. Oh S, Syn SY (2015) Motivations for sharing information and social support in social media: a comparative analysis of Facebook, twitter, delicious, YouTube, and Flickr. J Assoc Inf Sci Technol 66:2045–2060. https://doi.org/10.1002/asi.23320

65. Osatuyi B (2013) Information sharing on social media sites. Comput Hum Behav 29:2622–2631. https://doi.org/10.1016/j.chb.2013.07.001

66. Shi Z, Rui H, Whinston AB (2014) Content sharing in a social broadcasting environment: evidence from twitter. MIS Q 38:123–142. https://doi.org/10.25300/MISQ/2014/38.1.06

67. Tonteri L, Kosonen M, Ellonen HK, Tarkiainen A (2011) Antecedents of an experienced sense of virtual community. Comput Hum Behav 27:2215–2223. https://doi.org/10.1016/j.chb.2011.06.018

68. Sjöblom M, Hamari J (2017) Why do people watch others play video games? An empirical study on the motivations of Twitch users. Comput Hum Behav 75:985–996. https://doi.org/10.1016/j.chb.2016.10.019

69. Sjöblom M, Törhönen M, Hamari J, Macey J (2017) Content structure is king: an empirical study on gratifications, game genres and content type on Twitch. Comput Hum Behav 73:161–171. https://doi.org/10.1016/j.chb.2017.03.036

70. Furini M, Montangero M (2018) Sentiment analysis and twitter: a game proposal. Pers Ubiquit Comput 22:1–15. https://doi.org/10.1007/s00779-018-1142-5