# Mining Consumer Brand Relationship from Social Media Data: A Natural Language Processing Approach

Di Shang[1](✉), Zhenda Hu[2], and Zhaoxia Wang[3]

[1] University of North Florida, Jacksonville 32224, USA
d.shang@unf.edu
[2] Shanghai University of Finance and Economics, Shanghai 200433, China
[3] Singapore Management University, Singapore 178902, Singapore
zxwang@smu.edu.sg

**Abstract.** There is a rich collection of studies exploring different aspects of consumer brand relationship. Traditional approaches of questionnaires and analysis are based on measurements collected from a relatively small number of survey participants. With the advancements in natural language processing (NLP) techniques, opportunities exist for applying NLP techniques to discover consumer brand relationship from social media platforms that possess a large amount of data on consumer opinion and sentiment. In this study, we review consumer brand relationship analysis focusing on leveraging NLP and machine learning techniques to address some challenges associated with discovering customer brand relationship from social media data and propose a methodological framework for the approach. This study has implications for both academic research and practitioners as it presents an alternative way to investigate consumer brand relationship.

**Keywords:** NLP · Text mining · Consumer brand relationship · Machine learning

## 1 Introduction

The mass customization in many industries has caused increasing complexity of businesses operations in a global setting [1]. When a company operates over such a large geographical area, knowing consumer brand relationship is becoming more and more important for its business successes. Shimp and Madden were the first to introduce the concept of consumer brand relationship [2]. They defined consumer brand relationship (CBR) as "consumer forms a relation with consumption objects (products, brands, stores, etc.), which range from feelings of antipathy, to slight fondness, all the way up to the what would, in person-person relations, amounts to love." It is a relationship that consumers have with a product or a company which reflects how they think and feel about it. Aaker


X. Sun et al. (Eds.): ICAIS 2021, LNCS 12736, pp. 553–565, 2021.
https://doi.org/10.1007/978-3-030-78609-0_47

states that a business should think of a brand as a product, a person, an enterprise and a symbol to build the brand identity framework [3]. Most studies on consumer brand relationship rely on data collected from surveys in which participants directly report their perceptions in given dimensions. The survey studies often suffer from limited size of participants and prolonged time of preparing and conducting surveys.

With the advent of social media platforms, more and more consumers prefer to post their feedback, reviews and opinions of services and products online through social media platforms such as Twitter, Facebook and Google+ [4–6]. Such social media data provide a potentially rich source of information that can generate insights for better understanding consumer brand relationship. Collectively, such information may provide rich knowledge which may not be captured in traditional data collection methods such as administering a survey questionnaire. In addition, text mining technique and sentiment analysis based on social media data have been widely applied to brand authenticity, consumer brand relationship and engagement [7–9].

In this paper, we summarize the existing methods of applications of text mining and sentiment analysis to understand consumer brand relationship. Based on the previous research, we propose a methodological framework to assessing consumer brand relationship from social media as an alternative to traditional survey methods. Our framework allows researchers and practitioners to utilize existing survey items to develop training data for machine learning algorithms, incorporating large social media data by data augmentation and reducing the needs for direct survey inputs. We conduct a preliminary test of the approach with a secondary social media data set.

The rest of the paper is organized as follows: Sect. 2 discusses the existing consumer brand relationship research; Section 3 discussed related applications of data augmentation methods; Section 4 conducts a preliminary experiment on the proposed framework; finally, Sect. 5 concludes the paper by noting several limitations and suggesting potential improvements.

## 2 Related Work and Methods

### 2.1 Existing CBR Research Based on Surveys and Social Media

Previous studies have developed different dimensions that provide perspectives on consumer brand relationship and how relationships should be measured. Some suggest that the nature of brand relationship can be divided into three components: cognition, affection and conation [10], whereas Keller thought that brand relationships can be characterized in two dimensions of activity and intensity [11]. Barnes created a customer relationship index according to the interpersonal relationship measures [12]. Brand relationship index measures the relationship quality that a business has attained with consumers by 1) trust 2) commitment 3) awareness 4) attention 5) recognition and 6) interaction. The dimensions interact with each other, and sometimes are interchangeable. For instance, brand awareness establishes brand trust. Once consumers are bond to a brand, they are more likely to make repeat purchases, which then bridges the gap between trust and loyalty.

Social media has become a great way for businesses to engage with customers and build brands [13]. Social media platforms such as Facebook, LinkedIn, and Twitter give companies more ways to reach out to customers. More and more companies use Facebook Pages and Twitter handlers for brand communication and customer engagements [14]. Stephen and Galak describe two different types of social media involving a brand [15]. Earned social media is "referring to social media activity related to a brand that is not directly generated by the brand owner or its agents". In contrast, owned social media is "activity that is generated by the brand owner (or his/her agents) in social networking services that it can control". The applications of social media strategies help to increase such business values as raising brand awareness, boosting brand reputation, and enhancing brand loyalty [16]. Therefore, social media platforms should be used to create the illusion of a permanent brand presence in the informational environment of consumers. Brand loyalty level may be increased by an interactive, customized, and responsive way of social media use through networking and conversation with the consumers.

While the literature presents a rich collection of measurements of consumer brand relationship, all the studies rely on data collected from surveys. Consumers are asked to directly rate brand relationships on a numeric scale according to their perceptions in a given dimension. Some research suggests that rating brands via surveys can provide greater validity [17], and these are widely used in marketing practices. Though survey studies can obtain direct answers from customers or users, they suffer from some limitations. One limitation is its sample size. Researchers and industry practitioners have tried to increase sample sizes by sending surveys to hundreds and even thousands of customers, but they are still a small portion of the customer population. Another limitation is the time needed for conducting survey studies. Companies normally need to go through design, delivery, collection, and analysis processes, and it can easily take months for distributing and collecting surveys. Researchers have recognized the limitations of surveys such as the difficulty and expense in recruiting participants, collecting responses and ensuring truthful answers [18].

Social media and consumer review platforms have completely changed the way consumers communicate. Social media platforms help businesses gather richer, actionable insight about customer sentiment on their company, their brand, and specific products or services. Social media platforms have provided a valuable source for companies to collect information about a large number of customers in real-time. The abundant publicly available social media data and the advancements in NLP techniques motivate us to develop a flexible and automated means of estimating consumer brand relationship from social media data. The proliferation of social media use by both businesses and consumers offers a valuable data source to understand consumer perceptions. Text analysis of user-generated content (UGC) is a frequently used approach in the literature for mining consumer perceptions from social media data [19]. In marketing, researchers have employed text mining techniques on UGC to discover how product features or brands are perceptually clustered by consumers [20].

Many customer conversions on social media platforms reflect their opinions and attitude and thus can be used to measure consumer brand relationship. As shown in Table 1, well-established measurement items in survey studies can find their variants in social media data. The existing works, such as applications of NLP and text mining on social media [21–23] have laid a solid foundation for achieving the goals of this research. NLP and text mining have helped businesses and researchers to better know what is happening on the Internet, and it helps customer side as well.

**Table 1.** Examples of social media data related to brand relationship.

| Dimension | Source | Example |
|---|---|---|
| Commitment | Survey Item | I would like to recommend [brand] to my friends |
| | Social Media Data | [brand] is my number one favorite treat<br>I recommend these to anyone who wants to ... |
| Trust | Survey Item | This [brand] is reliable |
| | Social Media Data | I've been buying [brand] for 12 years and they never turn it down. Always a trustworthy company |
| Awareness | Survey Item | I know the product line of [brand] |
| | Social Media Data | [brand] just launched a new ... Compared with old models, ... |

## 2.2 Customer Brand Sentiment Analysis Based on Social Media

More and more marketing researchers have paid attention to social media and consumer review platforms in studying branding issues. Many companies keep abreast of customer sentiments to promote brands. Based on it, customer ratings and sentiment analysis can be used to quantify the overall positive and negative sentiments expressed online about a product or a brand [24], serving the same functions as ratings of survey respondents.

Specifically, some research utilized NLP and text mining for consumer brand sentiments by constructing domain lexicon. Mostafa used a random sample of 3516 tweets and an expert-predefined lexicon including around 6800 seed adjectives to evaluate consumers' sentiment towards famous brands such as Nokia and IBM, and results indicated a generally positive consumer sentiment towards several well-known brands [25]. Filipa et al. applied sentiment analysis to Yelp reviews of restaurants and built a specific aggregated dictionary for brand authenticity, consumer brand engagement and related constructs. Their findings revealed that online reviews could be an essential source of information for exploring sentimental attachment toward a focal object in brand authenticity and consumer brand engagement [26]. Bilro et al. also created a customer engagement dictionary based on WordNet 2.1 using Yelp customers' online reviews including 15,000 unique reviews of restaurants, hotels and nightlife entertainment. The findings indicated that the engagement cognitive processing dimension and hedonic experience had a significant effect on customers' review endeavor and customers seemed to be more engaged in positively advocating a company/brand than the contrary [27].

There are also some existing research in which the topics of social media users' representations and sentiments are identified to explore the relationship between customers and brands. Francesca and Alessandro utilized Emotional Text Mining to extract meaningful information for customer profiling and brand management. Specifically, the paper conducted the experiments on Twitter customer messages of a well-known sportswear brand to discover product preferences, representations, and sentiments. They proposed a bottom-up approach to classify unstructured data for the identification of social media users' representations and sentiments about a topic [7]. Byeongki et al. proposed an opportunity mining approach to identify product opportunities based on topic modeling and sentiment analysis of social media data. In their approach, latent product topics discussed by customers were identified in social media using topic modeling, thereby quantifying the importance of each product topic. Next, the satisfaction level of each product topic was evaluated using sentiment analysis. Finally, the opportunity value and improvement direction of each product topic from a customer-centered view were identified by an opportunity algorithm based on product topics' importance and satisfaction [9].

As for predicting customer sentiments towards brands, besides unsupervised methods, some supervised machine learning algorithms have also been adapted. For example, Hamid et al. constructed a procedure of brand authenticity sentiment analysis based on social media. Specifically, they established a framework using latent semantic analysis (LSA) and support vector machine (SVM) to predict both the brand authenticity dimensions and their sentiment polarity [8].

## 3 Data Augmentation Methods by Leveraging NLP Technology

Although the literature has built a solid ground of measurement items of consumer brand relationship, most survey studies develop or implement collections of only 20 to 30 items. While the number of items is sufficient to survey studies in which respondents are asked to directly rate brand relationships, it will not be feasible for NLP and machine learning approaches. NLP requires large databases of common phrases and sentences in a given language to understand or translate in natural language based on machine learning. In addition, the data about user comments on social networking sites are not always all available - the administrators often only allow positive comments and hide negative comments, which makes the dataset imbalanced [28]. The quality of NLP outcomes significantly depends on the availability of large corpora in a target domain. One way to increase the size of items for machine learning purpose is through data augmentation, and previous studies have concluded that text augmentation helps to further improve NLP model results [29].

Many data augmentation methods have been demonstrated to be effective. Zhang et al. found that one useful way to do text augmentation is replacing words or phrases with their synonyms [30]. In addition to the use of thesaurus, classic word embeddings, such as word2vec, GloVe and Fasttext, can be applied to perform similarity search. Fadaee et al. concluded that using contextualized word embeddings to replace target word outperforms static word embeddings [31]. Some other simple approaches have also been proved to be effective. They include inserting a synonym into a random position in the sentence, randomly choosing two words in the sentence and swapping their positions, and randomly removing a word in a sentence[1]. Similarly, Wei and Zou proposed easy data augmentation techniques for boosting performance on text classification tasks. The method includes four simple but powerful operations: synonym replacement, random insertion, random swap, and random deletion, which used only 50% of the available training set but achieved the same accuracy [32]. We apply some data augmentation techniques using NLPAug library to expand items adopted from [33], and some results from the augmentation are listed in Table 2.

**Table 2.** Results of data augmentation methods.

| Methods | Results |
|---|---|
| Synonym | I would like to **suggest** [brand] to my friends |
| Random insertion | **Post** I would like to recommend [brand] to my friends |
| Random swap | **my** would like to recommend [brand] to **I** friends |
| Random deletion | I would ~~like~~ to recommend [brand] to my friends |
| Original text | I would like to recommend [brand] to my friends |

In addition, NLP requires context in order to better understand the conversations of customers in social media. Different forms of context, such as cultural, places and events, have different effects on language understanding [34]. Corpora for machine learning should be representative and balanced with respect to the language that customers speak in a specific context. Incorporating contextual variables in NLP studies has shown to improve classification performance of social media analysis in that linguistic variations can be extreme [35]. Annotation can bring contextual variables into corpora, and annotated corpora can be used to train ML algorithms more efficiently and effectively [36]. An annotation development cycle normally follows model, annotate, train test, evaluate, and revise phases [37]. In order to provide statistically useful results, labelled data

[1] E. Ma. NLP Augmentation. https://github.com/makcedward/nlpaug, 2019.

on a sufficiently large corpus must be obtained [37]. Some examples of human annotated social media data based on measurement items of brand relationship from the literature are listed in Table 3. Consumer-brand related reviews are labeled with 1, and non consumer-brand related are labeled with 0.

**Table 3.** Examples of human annotated data.

| Text | Label |
|---|---|
| I have been buying [brand] for a few years and am very happy with it | 1 |
| This [brand] are my number one favorite treat. I recommend these to anyone who ... | 1 |
| This is a good product. My kids like them | 0 |
| Super fast shipping, will be purchasing them again! | 0 |

# 4 A Methodological Framework and Preliminary Testing

## 4.1 A Methodological Framework

Based on the discussion above, we propose a methodological framework for assessing brand relationship from social media data. Researchers and practitioners can use the framework to utilize well established survey items to conduct studies to discover consumer insights from social media data. As shown in Fig. 1, human annotation is often used to develop context-specific corpus for machine learning algorithms. Researchers can experiment with different annotation techniques to discover whichever best suit their research. Traditional survey instruments are well-established in the literature but usually have small data size. NLP augmentation techniques can be applied to increase the data size without compromising the data quality. Because of the diversified words and terms appearing in social media, word embedding techniques transform text data and provide a numerical representation that better captures the word meaning. Word vectors is a potent approach for quantitatively capturing word meaning. Embeddings generally represent geometrical encodings of words based on how frequently they appear together in a text corpus. The main idea is that a model can be trained on the context on each word, so similar words will have similar numerical representations [38].

## 4.2 Preliminary Testing

We use Amazon Fine Food Review dataset to test the proposed method to investigate consumer brand relationship from social media data. The dataset consists of 568454 reviews of 74258 fine foods from Amazon [39]. To test the

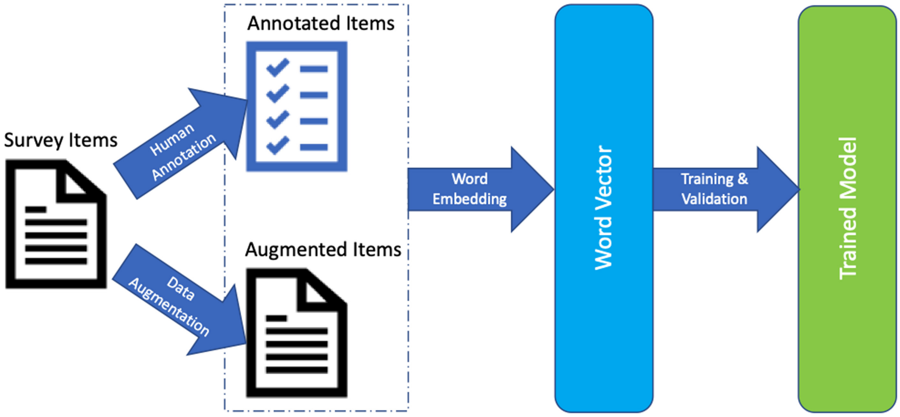


**Fig. 1.** Methodological framework.

generalizability of the proposed approach across brands, we retrieve five brands with the most reviews for our study, with a total of 3441 reviews. We adopt the well-established items from [33] (shown in Appendix) as the starting point to classify the review data. In this test, we focus on assessing consumer brand relationship in general, instead of examining individual dimensions.

**Table 4.** Summary of the deep learning model.

| Layer | Output shape |
|---|---|
| InputLayer | (None, 120) |
| Embedding | (None, 120, 300) |
| Conv1D | (None, 118, 128) |
| Conv1D | (None, 118, 128) |
| MaxPooling1D | (None, 29, 128) |
| MaxPooling1D | (None, 23, 128) |
| Concatenate | (None, 52, 128) |
| GRU | (None, 52, 128) |
| LSTM | (None, 128) |
| Dense | (None, 32) |
| Dense | (None, 1) |

NLP data augmentation is used to expand items using NLPaug library. Five graduate students worked as human annotators to mark a sample of 1000 reviews accordingly. The process resulted in 1000 labeled reviews, bringing up the training data to a total of 1480 labeled records. We applied FastText for word embedding because it uses n-gram characters as the smallest unit. In addition, FastText generates better word embeddings for rare words [40], which is very helpful for dealing with the rare words and typos which commonly appear in customer reviews. We used the data to train a deep network with an embedding layer (summarized in Table 4), and the model generated results with an average accuracy of 0.69, which show the feasibility of the proposed method.

## 5 Conclusions

In this paper, we summarize the existing methods of applications of text mining and sentiment analysis on consumer brand relationship and then introduce an approach for assessing consumer brand relationship from social media as an alternative to traditional survey methods. Our framework allows researchers and practitioners to utilize survey items to conduct consumer studies, incorporating large social media data and reducing the needs for direct survey inputs. In addition, using social media data, practitioners can monitor consumer brand relationship on social platforms and provide real-time recommendations to enhance the relationship.

The proposed method can be used to manage consumer relationships. Nowadays, customers expect to receive a customized and streamlined experience at all time. It is important to understand the preferences and purchasing habits of customers. The proposed method potentially benefits a business by making it simpler to track and investigate consumer relationship from social media. Businesses want to identify their customers quickly, address their main needs and also recommend extra services or products that may be of help to them. Understanding consumer relationship can help reduce the customer churn rate and identify which customers are most profitable.

We note several limitations with our experiment. First, our testing data is limited to brands of fine food. Future studies should examine a large range of brands. Second, consumers may express different opinions, concerns and sentiments across different social media platforms. Future studies should include other social media data, such as Twitter tweets, Yelp reviews and user forum threads. Furthermore, the noise and ambiguity of social media data present challenges to algorithmic solutions. Our proposed approach deserves further validation with a variety of social media data in different business contexts.

## A Appendix

Consumer Brand Relationship Measurement Items adopted from [33]

1. I would like to recommend brand X to my friends.
2. Image of brand X is fit for my taste.
3. In product Y, no other brands can replace brand X in my heart.
4. Image of brand X fits my current lifestyle.
5. If I buy product Y next time, I would like to buy brand X again.

6. If brand X is out of stock, I will go to another store to look for it instead of buying other brands.
7. I think highly of the prospect of brand X.
8. Although the price of brand X is a little bit higher than other brands, I would like to choose it.
9. I will not regret for choosing brand X.
10. I would like to buy other products of brand X.
11. The product of brand X satisfies my request for category Y very well.
12. I am satisfied with this [brand].
13. The [brand] has come up to my expectations.
14. This brand is close to an ideal [brand].
15. I pay attention to the news about company X.
16. I would like to visit the website of company X.
17. I would like to join the brand X club to communicate with more customers of brand X.
18. I know the requirement of typical customers of brand X for product Y.
19. The communication with brand X customers makes me feel intimate.
20. I would like to help brand X clients rather than other brands clients.
21. I would like to make friends with brand X customers rather than other brands customers.
22. I know the differences of product attributes (such as function, appearance, capability) between brand X and other brands.
23. I know the product line of brand X.
24. I know the business scope of company X.
25. I know the current prices of main brand X products.
26. I know the development history of company X.
27. I think that company X is familiar with the customers requirement for product Y.
28. I think that company X will deal with the feedback from customers.
29. I believe that company X will respect the customers' benefit.
30. I think that company X commitment to customers is credible.
31. This [brand] is reliable.
32. This is an honest [brand].
33. I trust this [brand].
34. I can recognize brand X only through its logo or advertising.
35. I can associate its advertising or logo with brand X's name.
36. I feel that I understand this [brand].
37. The [brand] and I are meant for each other.
38. This [brand] reveals a lot about my personality.
39. This [brand] plays a decisive role in my life.
40. I believe that this [brand] provides sufficient options to get in touch with other consumers/users of this [brand].
41. It is interesting to share experiences with other consumers/users of this [brand].
42. I use or would like to use the option to discuss with other consumers/users of this [brand].

43. I am of the view that this [brand] provides sufficient options to get in touch with employees of this [brand].
44. It is important to me being able to contact employees of this [brand].
45. I use or would like to use the option to discuss about [brand] with employees of this [brand].
46. I think that this [brand] provides sufficient options to get in touch with the [brand] producer through interactive online applications.
47. It is important to me being able to get in touch with the [brand] producer through interactive online applications.
48. I use or would like to use the option to get in touch with the [brand] producer through interactive online applications.

## References


1. Wang, Z., Joo, V., Tong, C., Xin, X., Chin, H.C.: Anomaly detection through enhanced sentiment analysis on social media data. In: 2014 IEEE 6th international conference on cloud computing technology and science, pp. 917–922. IEEE (2014)
2. Shimp, T.A., Madden, T.J.: Consumer-object relations: a conceptual framework based analogously on Sternberg's triangular theory of love. Adv. Consum. Res. **15**(1) (1988)
3. Aaker, D.A.: Building strong brands. Simon and Schuster, New York (2012)
4. Chou, W.Y.S., Hunt, Y.M., Beckjord, E.B., Moser, R.P., Hesse, B.W.: Social media use in the united states: implications for health communication. J. Med. Internet Res. **11**(4), e48 (2009)
5. Wang, Z., Chong, C.S., Lan, L., Yang, Y., Ho, S.B., Tong, J.C.: Fine-grained sentiment analysis of social media with emotion sensing. In: 2016 Future Technologies Conference (FTC), pp. 1361–1364. IEEE (2016)
6. Wang, Z., Ho, S.-B., Cambria, E.: A review of emotion sensing: categorization models and algorithms. Multimed. Tools Appl. **79**(47), 35553–35582 (2019). https://doi.org/10.1007/s11042-019-08328-z
7. Greco, F., Polli, A.: Emotional text mining: customer profiling in brand management. Int. J. Inf. Manag. **51**, 101934 (2020)
8. Shirdastian, H., Laroche, M., Richard, M.-O.: Using big data analytics to study brand authenticity sentiments: the case of starbucks on twitter. Int. J. Inf. Manag. **48**, 291–307 (2019)
9. Jeong, B., Yoon, J., Lee, J.-M.: Social media mining for product planning: a product opportunity mining approach based on topic modeling and sentiment analysis. Int. J. Inf. Manag. **48**, 280–290 (2019)
10. Reynolds, T.J., Gutman, J.: Laddering theory, method, analysis, and interpretation. J. Advert. Res. **28**(1), 11–31 (1988)
11. Keller, K.L.: Building customer-based brand equity. Mark. Manag. **10**(2), 14 (2001)
12. Fournier, S.: Secrets of customer relationship management: it's all about how you make them feel. J. Serv. Mark. **16**(7),700–703 (2002)
13. Liu, X., Shin, H., Burns, A.C.: Examining the impact of luxury brand's social media marketing on customer engagement: using big data analytics and natural language processing. J. Bus. Res. (2019)

14. Xie, K., Lee, Y.-J.: Social media and brand purchase: quantifying the effects of exposures to earned and owned social media activities in a two-stage decision making model. J. Manag. Inf. Syst. **32**(2), 204–238 (2015)
15. Stephen, A.T., Galak, J.: The effects of traditional and social earned media on sales: a study of a microlending marketplace. J. Mark. Res. **49**(5), 624–639 (2012)
16. Chierici, R., Del Bosco, B., Mazzucchelli, A., Chiacchierini, C.: Enhancing brand awareness, reputation and loyalty: the role of social media. Int. J. Bus. Manag. **14**, 153–160 (2019)
17. Bottomley, P.A., Doyle, J.R., Green, R.H.: Testing the reliability of weight elicitation methods: direct rating versus point allocation. J. Mark. Res. **37**(4), 508–513 (2000)
18. Steenkamp, J.-B., Van Trijp, H.: Attribute elicitation in marketing research: a comparison of three procedures. Mark. Lett. **8**(2), 153–165 (1997)
19. Fader, P.S., Winer, R.S.: Introduction to the special issue on the emergence and impact of user-generated content. Mark. Sci. **31**(3), 369–371 (2012)
20. Netzer, O., Feldman, R., Goldenberg, J., Fresko, M.: Mine your own business: market-structure surveillance through text mining. Mark. Sci. **31**(3), 521–543 (2012)
21. Coppersmith, G., Leary, R., Crutchley, P., Fine, A.: Natural language processing of social media as screening for suicide risk. Biomed. Inform. Insights **10**, 1178222618792860 (2018)
22. Press, V.G., Nyenhuis, S.M.: Do no harm: natural language processing of social media supports safety of aseptic allergen immunotherapy procedures. J.Allergy Clin. Immunol.**144**(1), 38–40 (2019)
23. Naaman, M.: Social multimedia: highlighting opportunities for search and mining of multimedia data in social media applications. Multimed. Tools Appl. **56**(1), 9–34 (2012)
24. Ludwig, S., et al.: More than words: the influence of affective content and linguistic style matches in online reviews on conversion rates. J. Mark. **77**(1), 87–103 (2013)
25. Mostafa, M.M.: More than words: social networks' text mining for consumer brand sentiments. Expert Syst. Appl. **40**(10), 4241–4251 (2013)
26. Rosado-Pinto, F., Loureiro, S.M.C., Bilro, R.G.: How brand authenticity and consumer brand engagement can be expressed in reviews: a text mining approach. J. Promot. Manag. **26**(4), 457–480 (2020)
27. Bilro, R.G., Loureiro, S.M.C., Guerreiro, J.: Exploring online customer engagement with hospitality products and its relationship with involvement, emotional states, experience and brand advocacy. J. Hosp. Mark. Manag. **28**(2), 147–171 (2019)
28. Luu, S.T., Van Nguyen, K., Nguyen, N.L.T.: Empirical study of text augmentation on social media text in vietnamese. arXiv preprint arXiv:2009.12319, 2020
29. Kobayashi, S.: Contextual augmentation: Data augmentation by words with paradigmatic relations. arXiv preprint arXiv:1805.06201, (2018)
30. Zhang, X., Zhao, J., Lecun, Y.: Character-level convolutional networks for text classification. Adv. Neural. Inf. Process. Syst. **649–657**, 2015 (2015)
31. Fadaee, M., Bisazza, A., Monz, C.: Data augmentation for low-resource neural machine translation. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (vol. 2: Short Papers), pp. 567–573 (2017)
32. Wei, J., Zou, K.: Eda: easy data augmentation techniques for boosting performance on text classification tasks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 6383–6389 (2019)

33. Zhou, Z.: How to measure brand relationship quality? Front. Bus. Res. China **1**(2), 300–318 (2007). https://doi.org/10.1007/s11782-007-0017-9
34. Garten, J., Kennedy, B., Sagae, K., Dehghani, M.: Measuring the importance of context when modeling language comprehension. Behav. Res. Methods **51**(2), 480–492 (2019). https://doi.org/10.3758/s13428-019-01200-w
35. Hovy, D., Søgaard, A.: Tagging performance correlates with author age. In: Proceedings of the 53rd annual meeting of the Association for Computational Linguistics and the 7th international joint conference on natural language processing (vol. 2: Short papers), pp. 483–488 (2015)
36. Neves, M., Ševa, J.: An extensive review of tools for manual annotation of documents. Briefings Bioinform. **22**(1), 146–163 (2021)
37. Saurı, R., Verhagen, M., Pustejovsky, J.: Annotating and recognizing event modality in text. In: Proceedings of 19th International FLAIRS Conference (2006)
38. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)
39. McAuley, J.J., Leskovec, J.: From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews. In: Proceedings of the 22nd international conference on World Wide Web, pp. 897–908 (2013)
40. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. Trans. Assoc. Comput. Linguist. **5**, 135–146 (2017)