9^{th} International Conference on Computational Social Science IC^{2}S^2
July 17-20, 2023, Copenhagen, Denmark

Analysis of psychographic indicators via LIWC and their impact on CTR for Instagram ads

Keywords: online advertising, Instagram, text mining, LIWC, CTR

Extended Abstract

In 2021, online advertising expenditures in Japan exceeded total advertising expenditures in traditional mass media (i.e., newspapers, magazines, radio, and TV) for the first time[1]. Display advertising accounted for 31.8% of these expenditures and was the main driver of growth[2]. As the online advertising market has grown, so has related research, particularly on CTR (click-through rate) prediction. Since the first report of CTR prediction for advertisements using logistic regression [1], various analytical models have been reported as useful for CTR prediction, but most CTR forecasts since the late 2010s have been based on deep learning models, such as the deep factorization machine (DeepFM [2]), proposed in 2017.

In psychology, traditional approaches of experimentation and surveys have been used to analyze individuals' behavioral tendencies and values, but with the development of an advanced information society, the use of big data to understand people's social lives through their online activity is becoming increasingly important[3]. Analysis of language content, which infers psychological tendencies from digitized text data, has been attracting a great deal of attention, and psychological research using text data, such as an analysis of Twitter content based on moral foundations theory [3], is increasingly being conducted. In online advertising, it is now possible to determine the linguistic content of appeals and link them to advertising effectiveness indicators, such as CTR. However, while research on advertising effectiveness prediction is active, the relationship between the linguistic content of appeals and advertising effectiveness indicators is still not clear.

In this paper, we calculated psychological indices of Instagram ad texts for health products and cosmetics using the J-LIWC2015 (the Japanese version [4] of the Linguistic Inquiry and Word Count[4]) [5]) and analyzed their correlations with CTR. The results show that negative appeals that arouse consumers' anxiety and sense of crisis were related to CTR and that when the product price was indicated in the advertisement, health products were negatively correlated with CTR while cosmetics were positively correlated.

We focused on two product categories—health products and cosmetics—and used data from Instagram ads served from January 1, 2020, to December 31, 2022, a summary of which is shown in Table 1. The product count represents the total number of products advertised on Instagram, with 48 in the health products category and 45 in the cosmetics category. Data were collected for 3,555 advertisements for health products and 3,270 for cosmetics, and the average CTR was calculated for each advertisement. The ad count is the total number of unique ads, which consist of combinations of images and text. Two types of text can be extracted from the advertisements: the main text, which is required for delivery with the image due to the format of Instagram advertisements, and the in-image text, which is contained within the body of the

[1] https://www.dentsu.co.jp/knowledge/ad_cost/2021/media.html
[2] https://www.dentsu.co.jp/news/item-cms/2022007-0309.pdf
[3] https://www.soumu.go.jp/johotsusintokei/whitepaper/h29.html
[4] http://www.liwc.net/

image. The in-image text was extracted using the OCR (optical character recognition) function of the Google Vision API, and the text data of the advertisements was quantified using the J-LIWC2015. The LIWC is a standard dictionary for text analysis that assigns to commonly used words a composite of linguistic categories pre-classified by experts (including both words with grammatical functions, such as auxiliary verbs and conjunctions, and words with substantive content, such as nouns and verbs) and categories related to psychological processes (including emotional, cognitive, and social processes) [4]. Using the LIWC, categories related to psychological processes embedded in the text data can be quantified and psychological states inferred. Prior to analysis of the text data, the J-LIWC2015 dictionary was applied to the source code published by Igarashi et al., and the necessary tagging was performed.

The correlation coefficients between the CTRs of the advertisements and the indices related to psychological process were calculated for the main text and the in-image text; health products are shown in Table 2 and cosmetics in Table 3. For health products, the main text showed weak positive correlations with "negemo" and "bio" and a weak negative correlation with "drives," while the in-image text had weak positive correlations with "affect" and "negemo" and weak negative correlations with "work" and "money." The use of biological terms such as "stomach," "weight loss," and "tiredness" when claiming efficacy may be a factor that increases CTR, when the advertisements with high "money" values had descriptions of product prices, which users may have perceived as advertisements and hesitated to click on them. For cosmetics, the main text showed weak positive correlations with "affect," "negemo," and "money" and a weak negative correlation with "percept," while the in-image text had a weak positive correlation with "bio" and a weak negative correlation with "relativ." "Percept" includes perceptual expressions about skin such as "bright," "smooth," and "glossy," and "Relative" includes expressions about "action," "space," and "time" (e.g., special sale until December 31), both of which may decrease CTR. On the other hand, the "bio" includes "skin" for body, and direct representation of skin may be a factor that increases CTR. Furthermore, there was a positive correlation with "money," which was contrary to the result for health products. For both, negative appeals that arouse consumers' anxiety and sense of crisis were found to be related to CTR. When prices were indicated in the ad text, health products were negatively correlated with CTR while cosmetics were positively correlated.

References

[1] Matthew Richardson, Ewa Dominowska, and Robert Ragno. Predicting clicks. In *Proceedings of the 16th international conference on World Wide Web*. ACM, May 2007.

[2] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. Deepfm: A factorization-machine based neural network for ctr prediction, 2017.

[3] Akiko Matsuo, Kazutoshi Sasahara, Yasuhiro Taguchi, and Minoru Karasawa. Development and validation of the japanese moral foundations dictionary. *PLOS ONE*, 14(3):e0213343, March 2019.

[4] Tasuku Igarashi, Shimpei Okuda, and Kazutoshi Sasahara. Development of the japanese version of the linguistic inquiry and word count dictionary 2015. *Frontiers in Psychology*, 13, March 2022.

[5] James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. The development and psychometric properties of liwc2015. 2015.

Table 1: Summary of data: "avg.CTR" is the average CTR per ad unit. avg.CTR for Health products (0.75%) is higher than avg.CTR for Cosmetics (0.61%).

Features	Details	Product category	
		Health products	Cosmetics
Product count	The total number of products advertised on Instagram	48	45
Ad count	The total number of unique ads	3,555	3,270
avg.CTR	Average CTR per ad	0.75%	0.61%

Table 2: Correlation coefficients between psychological category unit values and CTR for health products: Bolded values indicate values that are relatively correlated within a psychological category.

		main text	in-image text
affect		-0.029	**0.218**
	posemo	-0.130	-0.058
	negemo	**0.221**	**0.302**
personal concerns	home	-0.066	-0.024
	money	0.056	**-0.170**
	work	-0.044	**-0.190**
	death	-	0.006
	relig	-	-0.034
	leisure	-0.120	-0.013
social		-0.033	-0.116
cogproc		0.152	0.143
percept		0.027	-0.099
bio		**0.151**	0.022
drives		**-0.182**	0.115
relativ		0.013	-0.049
informal		-0.148	-0.003

Table 3: Correlation coefficients between psychological category unit values and CTR for cosmetics: Bolded values indicate values that are relatively correlated within a psychological category.

		main text	in-image text
affect		**0.226**	-0.025
	posemo	0.138	-0.073
	negemo	**0.254**	0.028
personal concerns	home	0.060	0.114
	money	**0.259**	-0.076
	work	-0.024	0.127
	death	-	-
	relig	0.126	-0.024
	leisure	0.078	-0.022
social		-0.009	0.002
cogproc		0.164	0.085
bio		**-0.308**	0.020
percept		0.028	**0.259**
drives		0.194	-0.028
relativ		0.106	**-0.192**
informal		0.094	-0.022