Do People Signal Their Socioeconomic Status on Social Media?−An Analysis of "Conspicuous Tweeting" by Machine Learning

Keywords: Socioeconomic Status, Cultural Capital, Tweets, Social Media, Signaling

Extended Abstract

The classical notion of "conspicuous consumption" (Veblen 1899) has been given a new life because people are posting their shopping, travel, or leisure activities on social media. Specifically, social media users may "signal" their socioeconomic status through the topics or vocabulary in the posts. This conjecture is weakly supported if their socioeconomic status is statistically predictive of their use of certain words or references to certain topics. Furthermore, it is strongly supported if their status is predicted by the content of their social media posts (Grimmer and Steward 2021).

Predicting the socioeconomic status, such as occupation, education, and income level, of social media users has been addressed using machine learning (e.g., He and Tsvetkova 2022; Sloan et al. 2015; Yo and Sasahara 2017). However, our study differs from existing studies in its main interest—status signaling conceived as a social phenomenon rather than accurately predicting socioeconomic status based on social media data. We integrated survey data measuring user attributes with social media posts to overcome the lack of sufficient information on user attributes in social media data (Stier et al. 2020), which were used to analyze their individual relationships. The details of this study are as follows (Figure 1 shows the framework).

First, we developed socioeconomic status scores for 500 occupational categories based on the Japanese version of the O-NET. Multiple web-based surveys were conducted using professional online research and crowdsourcing services. Approximately 100 respondents or workers per occupation category were employed residents of Japan, aged 20 years or older. These scores capture socially perceived levels of various aspects, such as prestige, power, income, education, culture, creativity, and social usefulness for each occupation. We focused on prestige, creativity, and social usefulness to understand the occupational structure in Japan because the occupational prestige score is highly correlated with scores other than creativity and social usefulness. Table 1 lists the top-ranking occupations in terms of scores.

Subsequently, we conducted a web-based survey of 5,196 Twitter users in the fall of 2021. The target population comprised Japanese residents aged 20 years or older who posted on Twitter twice or more per month. The respondents were asked about their socioeconomic status (occupation, personal and household income, and education) and demographics (age, gender, family structure, and area of residence). Approximately all occupations were converted into socioeconomic status scores that had already been developed. Additionally, participants were asked about their favorite leisure activities as proxies for cultural capital (Bourdieu 1979). More than half of the leisure activities were preferred by the respondents with more prestigious occupations than the rest. In addition, approximately half of those were less preferred by those with socially more useful occupations than the rest (Table 2). We confirmed that these activities represent the size of cultural capital, as hypothesized. Some leisure activities (creating art, taking artistic photos, playing video or computer games, and watching movies at home) are preferred by people whose occupations are more creative.

Twitter accounts were obtained with each respondent's permission, enabling us to collect past tweets of 3,930 respondents using the Twitter API. The survey and Twitter crawling data were integrated and analyzed in the following ways: one was predicting whether each Twitter user tweets about a brand based on their socioeconomic status, controlling for demographics and the number of words posted. Target brands were selected from a large-scale brand survey in Japan based on their level of brand recall. Logistic regression analysis showed that Twitter users with high occupational prestige tweeted about "cultural brands," such as newspapers and universities, more often than those with low prestige did, while users with high social usefulness tweeted about popular brands, such as Coca-Cola and P&G more often than those with low social usefulness did (Table 3).

Another direction of analysis was predicting socioeconomic status based on their tweet contents (usage of multiple words). If this prediction is accurate, people can infer a socioeconomic status of a Twitter user based solely on their tweets, suggesting that tweets may function well as status signals (Grimmer and Steward 2021). Otherwise, the tweets fail to signal the status despite such intentions of the Twitter user. After attempting alternative machine learning techniques, such as support vector machine (SVM) or random forest, L2 regularized logistic regression was applied because of its significantly predictive performance. The final model was selected using GridSearchCV. The predictions were considered successful only for age and gender because the recall of the model exceeded the actual proportions. However, the predictions failed for socioeconomic status variables. Table 4 summarizes the results.

Twitter users may tweet based on their status-specific preferences for brands; however, their status is not predictable from their tweets, implying that they do not signal their status. One possible interpretation of these contradictory results is that Twitter users tweet about certain brands that reflect their greater cultural capital; however, other disturbing factors make it challenging to detect status signals. Therefore, the prediction techniques, data preprocessing, or both must be refined. Another interpretation is as follows: Twitter users sometimes disclose status-based preferences in tweets; however, they do not intend to signal their status, and the undetectability of such signals is not surprising. It is plausible in Japanese society, where the tradition of social interaction across social strata exists (Ikegami 2005). Further analysis is required to study detailed mechanisms with which tweets are associated with socioeconomic status.

References

Bourdieu, P. (1979). *La Distinction, Critique Sociale du Jugement*. Paris, Les Editions de Minuit.

Grimmer, J., Roberts, M.E. and Stewart, B.M. (2021). Machine Learning for Social Science: An Agnostic Approach. Annual Review of Political Science, 24, 395–419.

He, Y. and Tsvetkova, M. (2022). A Method for Estimating Individual Socioeconomic Status of Twitter Users. *arXiv*:2203.11636.

Ikegami, E. (2005). *Bonds of Civility: Aesthetic Networks and the Political Origins of Japanese Culture*. Cambridge University Press.

Yo, T. and Sasahara, K. (2017). Inference of Personal Attributes from Tweets Using Machine Learning. *IEEE International Conference on Big Data,* 3168–3174.

Sloan, L., Morgan, J., Burnap, P. and Williams, M. (2015). Who Tweets? Deriving the Demographic Characteristics of Age, Occupation and Social Class from Twitter User Meta-Data. *PLOS ONE*, 10(3), e0115545.

Stier, S., Breuer, J., Siegers, P. and Thorson, K. (2020). Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field. *Social Science Computer Review*, 38(5), 503–516.

Veblen, T. (1899). *The Theory of the Leisure Class*. MacMillan Company.

Figures and Tables

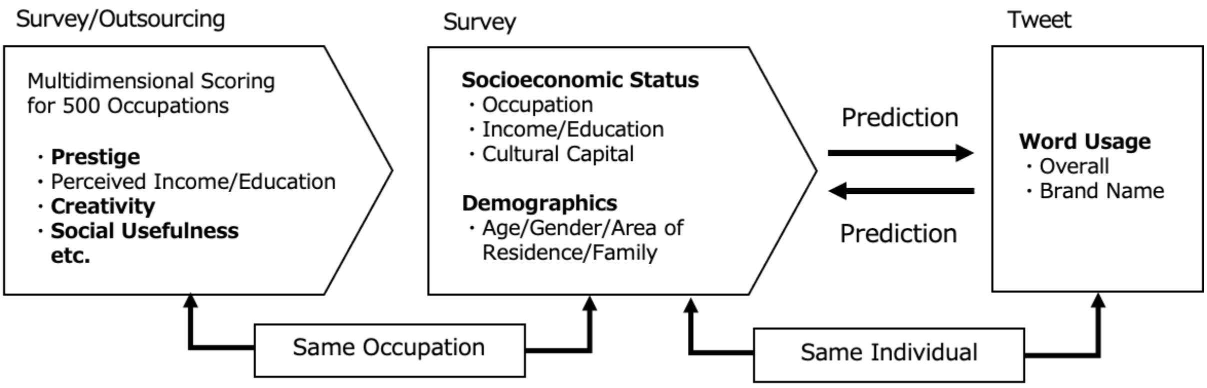

Figure 1. Framework of our study

Table 1. Top 10 occupations by representative socioeconomic scores

Rank	Prestigious Occupations	Mean Score	Creative Occupations	Mean Score	Socially Useful Occupations	Mean Score
1	Lawyer	4.81	Space Engineer	4.09	Nurse	4.42
2	Pilot	4.77	Video Editor	4.04	Pediatrician	4.31
3	Surgeon	4.76	Illustrator	4.03	Paramedic	4.30
4	Internist	4.73	Pastry Chef	3.97	Surgeon	4.29
5	Judge	4.70	CG Creator	3.97	Firefighter	4.29
6	Diplomat	4.69	Software Developer	3.95	Pharmacist	4.28
7	Pediatrician	4.61	Fashion Designer	3.95	Aircraft Mechanic	4.25
8	Space Engineer	4.60	Interior Designer	3.93	Garbage Collector	4.24
9	Other Physician	4.54	Actor/Actress	3.92	Pharmacologist	4.23
10	Prosecutor	4.54	Game Creator	3.90	Coast Guard Officer	4.22

Note) Scores are measured by 5-point scales from 1 to 5.

Table 2. Logistic regression of leisure activities on socioeconomic scores

Leisure Activities	*p*	Prestige			Creativity			Social Usefulness		
		coef.	s.e.		coef.	s.e.		coef.	s.e.	
Watching Movies at Cinema	36.3%	.155	.061	b	-.056	.095		-.359	.121	a
Reading for Pleasure	32.8%	.257	.063	c	.044	.097		-.668	.125	c
Playing Video Games at Home	30.7%	-.260	.066	a	.294	.102	a	-.032	.130	
Watching Movies at Home	29.1%	-.063	.065		.266	.101	a	-.081	.128	
Listening to Music Outdoor	27.5%	.130	.066	c	.004	.103		-.209	.132	
Listening to Music at Home	26.4%	.048	.067		.131	.104		-.341	.133	a
Doing Karaoke	21.2%	-.002	.074		.122	.114		-.078	.144	
Going to Pop Music Concerts	20.0%	.141	.073	c	-.110	.114		-.195	.145	
Going to Rock Music Concerts/Clubs	19.1%	.178	.075	b	.216	.114		-.285	.147	c
Going to Museums and Art Galleries	17.6%	.327	.079	a	.167	.120		-.791	.154	a
Watching Sports in the Stadium	17.6%	.355	.075	a	-.488	.121	a	-.353	.155	b
Training with Equipment	15.5%	.158	.080	b	-.025	.126		-.137	.162	
Jogging/Marathon Outdoors	14.8%	.311	.080	a	-.245	.129	c	-.080	.166	
Watching Performing Arts	13.9%	.332	.087	a	-.035	.132		-.724	.170	a
Going to Dance Music Concerts/Clubs	13.0%	.125	.088		.004	.135		-.565	.173	a
Playing Music Instruments or Singing	11.7%	.210	.091	b	.224	.139		-.518	.179	a
Not on the List	11.2%	-.215	.093	b	-.025	.148		.517	.186	a
Creating Artwork or Taking Artistic Photos	9.8%	.005	.101		.641	.150	a	-.500	.193	b
Going to Classical Music Concerts	9.1%	.439	.099	a	-.153	.155		-.553	.200	a
Going to Jazz Music Concerts/Clubs	7.9%	.420	.108	a	.369	.162		-.905	.213	a

Note) *p*: % of respondents performing each activity; coef.: estimated regression coefficients; s.e.: standard error; a, b, or c:1%, 5%, or 10% significance in the two-sided test, respectively. The results for the other variables were omitted from this table to save space.

Table 3. Logistic regression of tweeting brand names on user attributes

Brand	p	Prestige			Creativity			Social Usefulness		
		coef.	s.e.		coef.	s.e.		coef.	s.e.	
Yahoo!	20.2%	.241	.091	a	-.159	.139		-.146	.183	
Asahi Shimbun (Newspaper)	11.9%	.228	.115	a	-.203	.178		-.152	.236	
Nikkei Shimbun (Newspaper)	3.7%	.623	.176	a	-.151	.270		-.420	.366	
Tokyo University	1.4%	.786	.292	a	-.191	.436		-1.032	.587	
iPhone	7.0%	-.142	.144		.486	.218	b	.306	.280	
Starbucks	4.8%	.190	.170		-.611	.271	b	.339	.345	c
Apple	3.7%	-.265	.202		.592	.300	b	.125	.379	
Family Mart	15.5%	.087	.104		-.467	.165	a	.486	.210	b
Coca-Cola	10.3%	-.047	.125		-.306	.200		.785	.256	a
Kentucky Fried Chicken	9.9%	-.138	.133		-.235	.209		.526	.266	b
P&G	7.6%	-.182	.148		.262	.228		.630	.295	b
Twitter	17.1%	.023	.102		.018	.156		-.548	.205	a

Note) p: % of respondents tweet about the brand ten times or more; coef.: estimated regression coefficients; s.e.: standard errors; a, b, or c:1%, 5%, or 10% significance in the two-sided test, respectively. The results for the other variables were omitted from this table to save space. The coefficients mentioned above were insignificant for the other brands (Amazon, Facebook, Instagram, JOY, LIFE, LINE, LOTTE, McDonald, Mercari, Mos Burger, Nihon University, NTT DoCoMo, and Yomiuri Shinbun).

Table 4. Success/failure of prediction of user attributes based on tweet content

Successful Prediction	**Failed Prediction**
Gender (Male/Female)	Education (College Graduate/Others)
Age (40 years or older/Not)	Annual household income (10 million yen or higher/Not)
	Professional workers/Others
	Occupational prestige (continuous)
	Occupational creativity (continuous)