

Doshisha University Center for the Study of the Creative Economy

Discussion Paper Series No. 2014-01

Determinants of Movie Review Ratings

-New method by using Big Data -

Tadashi Yagi (Doshisha University)

Seiya Murata (Doshisha University)



Discussion Paper Series

Determinants of Movie Review Ratings

-New method by using Big Data -

Tadashi Yagi

(Doshisha University)

Seiya Murata

(Doshisha University)

Abstract

We present a method for analyzing the determinants of ratings in reviewing creative works by examining the case of movies in Japan. In Japan, the Yahoo Movie Review website presents viewer ratings, comments, the number of followers, and evaluations by followers. We develop a system for extracting information from this website and analyzing keywords in comments for positive and negative emotional expressions by using the “IBM SPSS Text Analytics for Survey” text-mining software. This allows conversion of text data to numerical data, and application of principal factor analysis and regression analysis. We generate variables representing evaluations of “plot,” “cultural value,” “main theme,” and “setting” of each movie, and analyze the sign, positive or negative, of each variable’s coefficient in a rating by regression analysis. We present a case showing positive evaluations for plot, cultural value, and setting, but negative evaluation for theme.

1. Introduction

Various Internet websites allow reviewing movies and performances. These websites can be regarded as a resource for improving the quality of creative works so long as the information presented can be statistically analyzed at a reasonable cost. In this context, automatic review mining and summarization has been gaining attention from various fields, such as product and service businesses.

Zhuang, et. al. (2006) focus on movie reviews, and find that reviews have unique characteristics. Specifically, movie reviews address not only movie elements such as screenplay, but also movie-related roles such as directors, screenwriters, and actors. They decompose review mining and summarization into 1) identifying characteristic and opinion words, 2) determining the class of feature word and its opinion polarity, 3) for each feature word, identifying relevant opinion words and obtaining valid feature opinion pairs, and 4) summarizing of the discovered information.

Chaovalit, Pimwadee, and Lina Zhou (2005) investigated movie review mining by two approaches: machine learning and semantic orientation. Mohammadi, Gelareh, Sagae, Vinciarelli, and Morency (2013) explored the effect of multi-modality and perceived personality on persuasiveness of social multimedia content by using movie review clips from YouTube. Their interest is similar to ours in that they tried to clarify factors affecting degree of persuasion.

We present a method for analyzing the determinants of ratings in reviewing creative works by using the case of movies in Japan. In Japan, the Yahoo Movie Review website presents viewer ratings, comments, the number of followers, and evaluation by followers. As a first step, we developed a system of automatically extracting information from the website, and generating an Excel-format database. This system was developed in the PHP programming language after analyzing the HTML format of pages on the website. In the second step, we developed a system for extracting information from this website and analyzing keywords in comments for positive and negative emotional expressions by using the “IBM SPSS Text Analytics for Survey” text-mining software. This allows conversion of text data to numerical data, and subsequent application of principal factor analysis and regression analysis.

We then generated variables representing evaluations of “plot,” “cultural value,” “main theme,” and “setting” of each movie, and analyzed the directional effect of each variable on rating by regression analysis. This paper presents a case showing positive relations for plot, cultural value, and setting, but a negative relation for theme. More elaborate development of the analysis method is expected to facilitate detailed clarification of factors that stimulate viewers’ interest.

2. Method

2.1 Characteristics of the movie review site

In Japan, “Yahoo Movie Review” is a well-known movie review site. The characteristics of this site are summarized as follows.

1. Free-description-type comments

Reviews include text comments containing various emotional expressions that can be linked with various types of words, including nouns.

2. Evaluation scale

Reviewers evaluate movies on a scale from 1 (lowest) to 5 (highest).

3. Degree of review usefulness evaluations

Reviews themselves can be graded by readers, especially regarding review usefulness.

4. Posting date and time information is available

Posting date and time information are used in adjusting the number of followers, because the number increases as time passes.

We use reviews of the movie “Ask This of Rikyu (*Rikyu ni Tazuneyo*)” for analysis. The address of the review site is <http://info.movies.yahoo.co.jp/detail/tymv/id345694/>. This is a historical drama based on a novel by Kenichi Yamamoto, for which he won the Naoki Prize. The main theme of the movie is the “aesthetic” pursued by an historic tea master Sen no Rikyu. A controversial plot point is that Rikyu was inspired by the aesthetics of a Korean woman who was kidnapped by the Japanese. It

is a historical fact that Rikyu loved tea cups and various items related to tea ceremony imported from Korea. However, many Japanese people are reluctant to accept this fact, and believe that the tea ceremony originated in Japan. For this reason, both extremely positive and extremely negative reviews are presented. An example negative review follows.

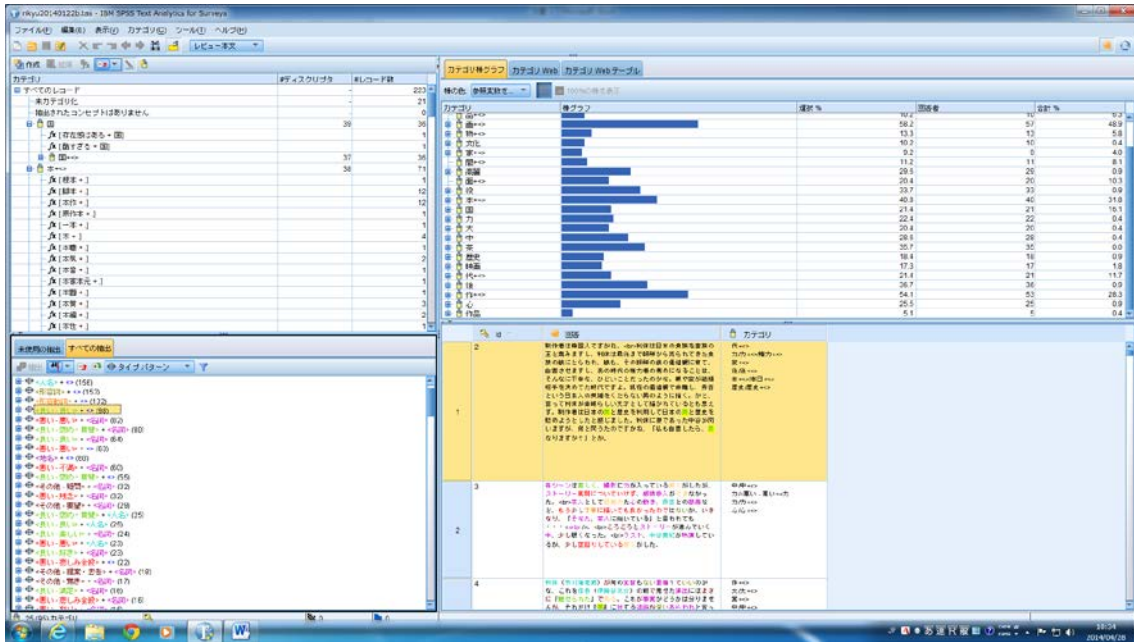
The presented origin of the tea ceremony is unacceptable, delivered in a manner like a Korean drama, which is quite sloppy. In the movie, Rikyu explores aesthetics, but its source is a tragic love of his youth. The movie proclaims that this aesthetic, which originates in Korean culture, overwhelms the power of authority, and thus cannot be transmitted during Rikyu's life. Rikyu committed ritual suicide to protest the actions of Hideyoshi Toyotomi [a preeminent daimyo, warrior, general, and politician of the Sengoku period]. The movie suggests that the true reason for his suicide is a mystery. I think Rikyu acts as a foil to the complexity of Hideyoshi because of his origin as a farmer, and he could not understand aesthetics fundamentally. This invoked hatred between Hideyoshi and Rikyu. Rikyu said "I obey only aesthetics," and it is natural that authority could not tolerate these words. Ebizo [a famous kabuki actor] exhibited a certain aura while playing Rikyu, but I think his charm is not as fully exhibited as it is in kabuki.

We investigate factors that determine the final evaluation of reviews by examining the review text and evaluation points given by reviewers. This task is important because impressions of reviews are biased by highly emotional words, making it hard to examine the actual factors that determine the evaluation. For example, words such as "disappointed," "nonsensical," "absurd," "below the level of B-movie quality," "regret seeing it," and "contempt for Japanese aesthetics and history" are observed in the reviews. Without analysis, such descriptions may make readers believe that the movie is not worth viewing. It is not obvious, however, how assigned movie ratings and ratings of the reviews themselves are affected by these words. Analysis allows examining the actual factors that affect grading and clarifies the true functioning of movie review sites.

2.2 Text mining

We used "IBM Text Analytic Survey" as text mining software. The first step was to extract words, and the second was to categorize those words. When categorizing words, emotional words were used to link nouns and verbs. By selecting the type of category, we can depict linkage among words. Here, we selected positive emotional words and extracted the related words.

Figure 1. Screen of IBM Text Analytic Survey



As is shown in Figure 1, the text mining software relates words with emotional words such as “good,” “bad,” “praise,” and “grief.” The number of words relating to these emotional types is listed in the bottom-left window of the software. By selecting, for example, “good” and “praise,” we can produce a bar chart of the frequency of related words in the upper right window. We then exported this output as an Excel file, combined it with the original Excel file, and produced the SPSS file.

2.3 Extracting factors that determine evaluation grading

The next task is to extract factors that determine evaluation grading. We apply principal factor analysis to reduce the number of variables and classify the words. As a result of the analysis, we extracted four factors for words associated with positive emotional words.

Table 1. Sum of explained variance in principal factor analysis for words associated with positive emotional words

Factor	Initial eigenvalue			Sum of squares of loadings after extraction			Sum of squares of loadings after rotation
	Sum	% of variance	Cumulative % of variance	Sum	% of variance	Cumulative % of variance	Cumulative % of variance
1	3.792	13.077	13.077	3.059	10.548	10.548	5.552
2	2.020	6.965	20.042	1.243	4.287	14.835	10.916

3	1.618	5.581	25.622	.877	3.025	17.860	15.450
4	1.490	5.138	30.760	.729	2.513	20.374	19.740

Table 2. Structural matrix of principal factor analysis for words associated with positive emotional words.

	Plot	Tea items	Culture	Tea ceremony
Created + <>	.545	.026	.021	.096
Mind / heart + <>	.238	.141	.073	.214
Tea / tea + <> / tea + <>	.108	.373	-.028	.190
Generation + <>	.122	.169	-.028	.412
Power / force + <> / power + <>	.024	.026	.128	.360
Movie + - <good fun> Movies /	-.027	-.067	.182	-.013
Book + <>	.289	.202	.084	.090
Medium / medium + <>	.224	.173	.084	.377
Surface + <>	.259	.349	.137	.151
Large / large + <>	.253	.517	-.002	.212
This + <> / today + <> / Japanese culture + <>	-.050	-.064	.162	.079
Tea / tea + <> / tea ceremony + <>	.019	.059	.235	.321
Power / force + <>	.286	.077	.494	-.079
Created + <> / Movies + <>	.490	.059	.055	.214
House + <>	-.045	.241	-.046	.242
Between + <>	.031	.315	-.015	.041
Culture / Culture + <>	-.026	.056	.481	-.013
Goryeo / Koryo + <>	.153	.446	.198	.074
After / rear + <>	.277	.225	-.014	.173
Goods + <>	.145	.242	.010	-.026
Country / country + <>	-.061	.070	.463	.153
Role / role + <>	.591	.198	.068	-.102
Road + <>	.104	-.036	.015	.465
Image + <>	.336	.171	-.221	.154

This + <> / today + <>	-.014	-.010	.294	.193
Field + <>	.077	.076	.192	.099
History / History + <>	.038	-.258	.217	.202
Thing + <>	.136	.515	.149	-.112
Real + <>	.105	.070	.404	.057

Tables 3 and 4 present the results of principal factor analysis for words associated with negative emotional words.

Table 3. Sum of explained variance in principal factor analysis for words associated with negative emotional words.

Factor	Initial eigenvalue			Sum of squares of loadings after extraction			Sum of squares of loadings after rotation
	Sum	% of variance	Cumulative % of variance	Sum	% of variance	Cumulative % of variance	Cumulative % of variance
1	2.867	11.027	11.027	2.326	8.947	8.947	4.337
2	1.782	6.855	17.883	1.210	4.653	13.600	8.563
3	1.688	6.490	24.373	1.072	4.123	17.723	12.743
4	1.628	6.260	30.633	1.051	4.043	21.766	16.325
5	1.222	4.698	50.462	.622	2.392	32.653	29.480

Table 4. Structural matrix of principal factor analysis for words associated with negative emotional words.

Factors	Rikyu	Spiritual	Nationalism	History	Korea
Movie + - <bad bad> Movies /	-.040	-.035	-.027	.132	-.035
Movie + - <good good> Movies /	.001	.020	.011	-.067	-.031
<Good - praise, praise> Movies / Movie +	.354	-.047	-.009	-.026	.143
Movie + - <good fun> Movies /	-.109	.267	-.071	.121	.063

<Good - praise, praise> Movies / Movies +	-.005	.077	.085	-.038	.041
Role / role + <>	.299	.023	-.003	.047	.055
History + - <bad bad> History /	-.116	-.064	-.144	.537	.002
Goryeo / Koryo + <> / Goryeo woman + <>	-.052	-.110	.395	.020	.142
History / History + <>	.163	.039	.128	.648	-.101
Goryeo / Koryo + <>	.184	.108	.272	-.046	.776
Country / country + <>	-.161	.054	.555	.029	.150
Mind / heart + <>	.109	.712	.037	.001	.010
Mind / heart + <> / suicide + <>	-.024	.270	-.066	-.020	.244
Field + <>	.232	-.124	.219	.011	-.022
Surface + <>	.186	.271	.200	-.058	.104
House + <>	.066	-.001	.034	.002	.106
Between + <>	.054	.079	-.088	-.053	.024
Tea / tea + <> / tea + <>	.095	.033	.180	-.044	.045
Tea / tea + <> / tea ceremony + <>	.034	.017	.492	-.004	-.060
Tea / tea + <>	.273	.468	.068	.059	.026
Real + <>	-.016	.108	.209	.392	.083
Road + <>	.205	.238	.254	.093	-.153
Image + <>	.650	.089	-.032	-.033	-.019
Goods + <>	.025	.057	-.006	.054	.059
This + <> / real + <>	.349	.082	-.056	.027	.038
Book + <>	.130	.054	.068	.004	-.068

3. Regression analysis on the determination of factors for grading

We conducted regression analysis with the factors derived by principal factor analysis on the set of words associated with either positive or negative emotional words. Table 5 and Figure 2 show the results for grading evaluation. The standardized coefficients are comparable in size because the variable unit is normalized. Larger values of the standardized coefficient imply that the effect on grading evaluation is stronger. The results show that “Tea Ceremony” was associated with positive emotional words, and had the strongest positive effect on evaluation grading. “Rikyu” was

associated with negative emotional words, and had the second strongest positive effect on evaluation grading. “Nationalism” and “history” associated with negative emotional words, and decreased evaluation grading. “Korea” was associated with negative emotional words, and increased evaluation grading.

Table 5. Regression result for factor determination for grading evaluation

	Coefficient		Standardized coefficient	<i>t</i>	<i>p</i> -value
	B	Standard errors			
(Constant)	2.583	0.091		28.254	0
Tea Items (positive)	-0.299	0.222	-0.155	-1.348	0.179
Tea Ceremony (positive)	0.655	0.205	0.329	3.198	0.002
Rikyu (negative)	0.399	0.168	0.207	2.378	0.018
Spiritual (negative)	0.223	0.141	0.119	1.58	0.116
Nationalism (negative)	-0.379	0.214	-0.194	-1.767	0.079
History (negative)	-0.382	0.157	-0.197	-2.434	0.016
Korea (negative)	0.266	0.14	0.148	1.909	0.058

Adjusted R²: 0.183

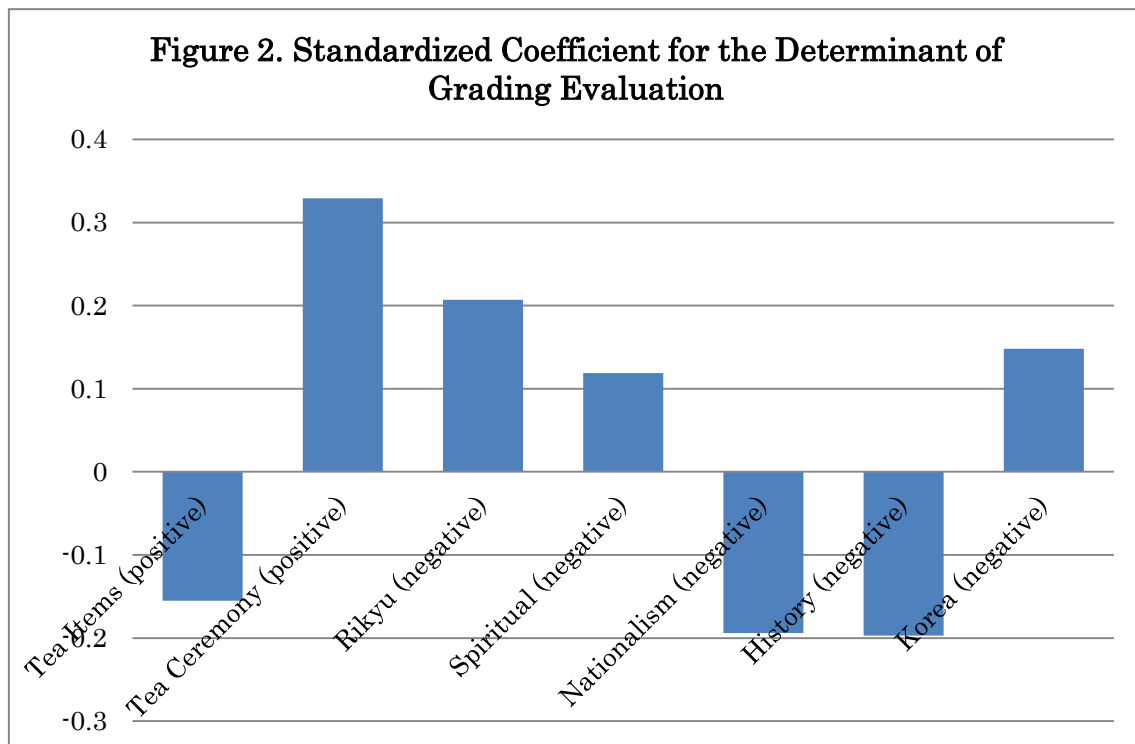
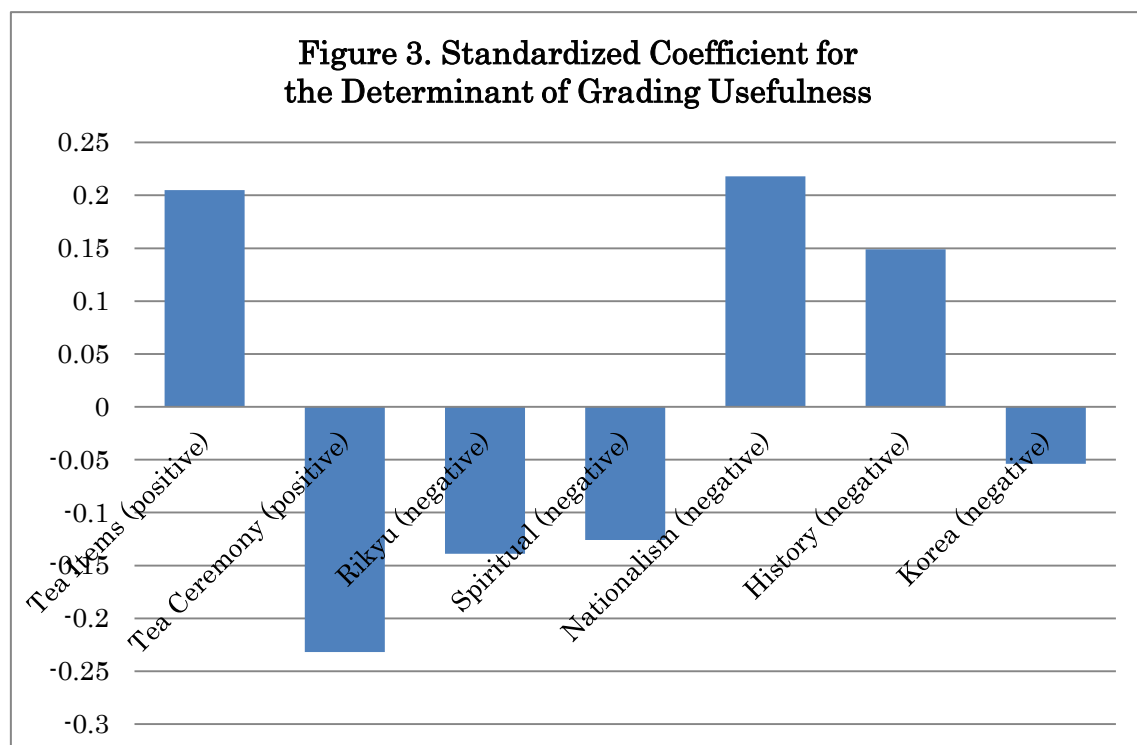


Table 6 and Figure 3 show the results for grading usefulness. “Tea items” was associated with positive emotional words and had a positive effect on grading usefulness. “Tea ceremony” was associated with positive emotional words, but had the strongest negative effect on grading usefulness. “Nationalism” and “history” were associated with negative emotional words and had a strong positive effect on grading usefulness.

Table 6. Regression results for factor determination for grading evaluation

	Coefficient		Standardized coefficient	<i>t</i>	<i>p</i> -value
	B	Standard			
(Constant)	38.094	3.108		12.258	0
Tea Items (positive)	12.917	7.54	0.205	1.713	0.088
Tea Ceremony (positive)	-15.114	6.963	-0.232	-2.171	0.031
Rikyu (negative)	-8.756	5.702	-0.139	-1.535	0.126
Spiritual (negative)	-7.75	4.809	-0.126	-1.612	0.109
Nationalism (negative)	13.92	7.282	0.218	1.911	0.057
History (negative)	9.447	5.329	0.149	1.773	0.078
Korea (negative)	-3.197	4.743	-0.054	-0.674	0.501

Adjusted R²: 0.116



4. Relation between evaluation grade and sales

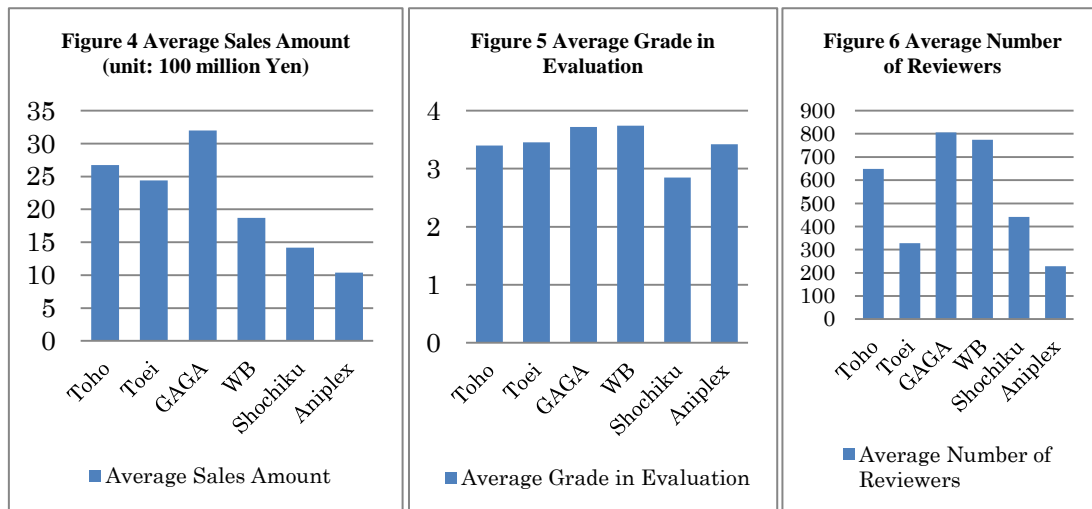
To understand the role of movie reviews in more detail, we examined factor determinants that explain sales. Data were taken from the website of the Motion Picture Producers Association of Japan (<http://www.eiren.org/toukei/index.html>). Table 7 shows the results of regression analysis on sales determinants. The number of evaluators is strongly associated with sales, but evaluation rankings had no explanatory power. Toho and Toei are the two largest Japanese movie companies, but as Fig. 4 shows, GAGA achieves the largest average sales per movie.

These results suggest that high evaluations do not guarantee the commercial success of a movie. Commercial success is, however, reflected in the number of reviewers. It is interesting that Toei films have higher average evaluations than those from Toho, but the average sales and number of reviewers are smaller. In particular, compared with Toho films, Toei films had far fewer reviewers. One possible explanation is that Toho targets mass-market consumers, but Toei produces movies targeting film connoisseurs.

Table 7. Regression analysis on sales determinants

	Coefficient		Standardized coefficient	<i>t</i>	<i>p</i> -value
	B	Standard			
(constant)	8.738	12.235		.714	.481
Grade in evaluation	-1.371	3.321	-.039	-.413	.683
Number of reviewers	.020	.002	.842	8.753	.000
Toho_dummy	9.341	5.196	.225	1.798	.083
Toei_dummy	13.690	6.538	.257	2.094	.045
GAGA_dummy	11.842	12.240	.098	.967	.341

Adjusted $R^2=0.693$



5. Conclusion

The results of regression analysis show the merits and the limitations of this analytical procedure. This analysis allows separation of factors influencing reviews and those that do not. Without this kind of analysis, it is difficult to distinguish whether factors associated with emotional words actually affect grading, even when the emotional words give a strong impression to review readers. One limitation of this analysis is ambiguity in interpreting the results. “Tea ceremony” was associated with positive emotional words and had a positive effect on grading evaluation. The result itself gives limited interpretable information. To interpret the results, it is therefore necessary to defer to the original comments. Alongside such follow-up tasks, the information given by this kind of analysis provides us with precious information regarding movie reviews.

Finally, we would like to discuss utilization of review sites by movie producers. We showed that commercial success is not guaranteed by review evaluations. However, this does not imply that the grade does not reflect movie quality. The method we propose allows judging the true factors in determining evaluation grading and gives us detailed information on the relation between quality and evaluation. In the movie market there exists a stratification of viewer quality, where mass-market viewers can be considered of medium or low quality. Commercial success in the mass market does not necessarily guarantee long-term success in the movie business without improvement to movie quality.

References

- Chaovalit, Pimwadee and Lina Zhou (2005), “Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches,” Proceedings of the 38th Hawaii International Conference on System Sciences – 2005.
- Lara, Frankie (2012), “Creativity and the Cinema Movie Review: It’s Kind of a Funny Story,” *Journal of Creativity in Mental Health*, 7:304–307.
- Mohammadi, Gelareh, Kenji Sagae, Alessandro Vinciarelli, Louis-Philippe Morency (2013), “Who is persuasive?: the role of perceived personality and communication modality in social multimedia,” *ICMI '13 Proceedings of the 15th ACM on International conference on multimodal interaction*, Pages 19-26, ACM New York.
- Wijdicks, F.M. Eelco (2011) “Movie Review: Do Neurologists Need To Know Jack?” *Neurology Today*, January 6, 2011.
- Zhuang, Li, Feng Jing, and Xiao-Yan Zhu (2006), “Movie Review Mining and Summarization,” *CIKM'06*, November 5–11, 2006.