Relationship Between Audience Engagement On Social Media And Broadcast Media Ratings

Sangun Park

Juyoung Kang

Youngsok Bang

Kunsoo Han

RELATIONSHIP BETWEEN AUDIENCE ENGAGEMENT ON SOCIAL MEDIA AND BROADCAST MEDIA RATINGS

Sangun Park, Kyonggi University, South Korea, supark@kgu.ac.kr
Juyoung Kang, Ajou University, South Korea, jykang@ajou.ac.kr
Youngsok Bang, The Chinese University of Hong Kong, bangyoungsok@baf.cuhk.edu.hk
Kunsoo Han, McGill University, kunsoo.han@mcgill.ca

ABSTRACT

People often share their opinions or impressions about TV shows (e.g., dramas) with other viewers through social media such as personal blogs and Twitter. As such, broadcast media, especially TV, lead to audience engagement on social media. Moreover, the audience engagement, in turn, impacts broadcast media ratings. Social TV analyzes audience's TV-related social media behaviors and tries to use the behaviors in marketing activities such as advertisement; however, this is purely based on the quantity of engagement in social media. In this study, we analyze the subjects of the audience engagement on social media about specific TV dramas through topic modeling, and examines the relationship between changes in the topics and viewer ratings of the TV dramas.

Keywords: Social media, broadcast media, topic modeling, topic trend, VAR.

INTRODUCTION

Social media is a hot research topic nowadays. The responses in social media helps prediction and estimation in various areas such as election [6] and marketing campaign [10]. Recently, the relationship between social media and other types of media is drawing attention. The purpose of this paper is analyzing the relationship between audience engagement on social media and broadcast media ratings. It looks somewhat complicated but the basic idea is very simple. People watches TV and talks about it on social media, so we can guess that the viewing TV shows will affect conversation about the TV shows on social media. Moreover, we can guess that the conversation will also affect viewing again, which means someone who participated in the conversation but did not watch the shows can decide to watch the shows. However, this idea is not a very new one. For example, Social TV is a study about television-related social behavior [2]. They tried to support communication and social interaction around TV with new applications [3]. Some studies about social TV suggested measures on social media activities tied to specific TV broadcasts. Bluefin [4] is an example of such social TV that focused on the intersection of TV and social media. The company tried to provide new insight into audience engagement and suggested two types of metrics to measure the audience response to television by using conversations in social media. One metric is measured the number of commenters for episodes of a drama in the response level. The other metric is about response share that is the percentage of the share of audience response in social media.

Social TV was the hottest area once, but an article of GIGAOM at 2014 showed that most of the social TV efforts failed [6]. It presented some evidences for the failure. Yahoo announced that it would shut down Intonow and i.TV also discontinued GetGlue. In addition, most of the social TV startups such as Matcha, Tunerfish and Screentribe failed to make TV social. However, we don't think it means that the insight from social media is useless. They focused on providing social media services for TV. But, there were already Twitter and Facebook that are elephants in social media services. The insight from social media response is still very important. It is just time to move on and try something new [11].

We believe that one of those "something new" is a semantic analysis of social media conversations. Most of the metrics suggested by social TV are quantitative analysis which counts the number of certain keywords in social media conversations. However, we are interested in the topics discussed about TV in social media. Once we extract topics from social buzz with text mining tools such as topic modeling [13], we also can analyze the relationship between the semantics of social media topics and viewer ratings of a certain TV show. In other words, we can find how they affect each other.

Therefore, the fist research question of our study is what topics about TV dramas are discussed in social media. We can use topic modeling which is one of popular text mining techniques in this semantic analysis [8]. The second question is if the topics and viewer ratings have certain trends over time. Topic trend analysis will be used for this question by using the results of topic modeling. The last question is that there is a relationship between the topics and the viewer ratings. There could be various relationships between topics or between a topic and viewer ratings. The vector autoregression (VAR) which is an econometric model for the analysis of the linear interdependencies among multiple time series will be used for the question [12].

RESEARCH METHODOLOGY

Why Dramas Are Important?

There is a report about the economic ripple effect of "My Love from the Star" which is a very popular and successful Korean drama in Asia [5]. It estimated the overall sales including advertisement sales, oversea sales, related products and tourist as $478 million. Moreover, domestic production inducement effect was estimated as $910 million and value-added inducement effect was estimated as $374 million. It is a well-known story that Netflix used big data analysis in the design of the famous drama "House of Cards". The article about the background and detailed story of which title is "Giving Viewers What They Want" [1] insists that big data analysis required for finding tastes and requirements of audience is very important in designing a drama, and it is a formal

trend now.

Selection of Social Media and Dramas

We chose five Korean dramas based on the rank of each drama category [7]. Table 1 shows a brief summary of the dramas. We collected data including viewer ratings about dramas from Naver.com(http://search.naver.com/search.naver?where=nexearch& query=%EB%B3%84%EA%B7%B8%EB%8C%80&sm=top_hty&fbm=1&ie=utf8) which is the representative portal in South Korea. Naver.com provides viewer ratings of each episode of dramas provided by Nielson Korea(http://www.agbnielsen.co.kr/). Viewer ratings, which is called Nielsen ratings (https://en.wikipedia.org/wiki/Nielsen_ratings), are the audience measurement systems to determine the audience size. It is required to give attention to the number of episodes. Some of dramas have a quite small number of episodes for statistical analysis. Moreover, those were not broadcasted every day of the week. It means that the time-series data are irregular. The two things became problems in our analysis.

Table 1.Summary of Selected Dramas

Company	Title	Period	Number of episodes	Broadcasting days (Category)	Rank	Viewer Rating (Avg.)
KBS	Wang Family	2013.08.31 ~2014.02.16	50	Weekend	1	33%
MBC	Empress Qi	2013.10.28 ~2014.04.29	51	Mon, Tues	1	21.9%
SBS	Dr. Stranger	2014.05.05 ~2014.07.08	20	Mon, Tues	2	11.7%
SBS	My Love from the Star	2013.12.18 ~2014.02.27	21	Wed, Thurs	1	24%
tvN	Misaeng	2014.10.17 ~2014.12.20	20	Fri, Sat	1	5.4%

The next step was choosing a social media service. There is a report about the social media usage in South Korea [9]. Facebook was the number one social media service with 77.4% of usage share, but it is very difficult to crawl the contents from Facebook because it just provides search API on the connections between users instead of general search on every walls. Therefore, we chose the second social media type, Blogs of which market share is 30% and it is easy to search with keywords. Moreover, the contents are long enough for text mining. Among the various Blog services in Korea, we chose NAVER Blog because its market share is 71.2% in the Blog market area.

Analysis Framework

Our analysis framework consists of four steps which are Blogs crawling, topic modeling, topic trend analysis and VAR analysis. The first step is crawling documents from the NAVER blogs. This step consists of getting HTML documents from the blogs and parsing them. After that, we extracted the original form of nouns through morphological analysis. The second step is topic modeling with the crawled documents. We used Mallet which is a well-known topic modeling tool. In this step, we analyzed the topics and named those. The third step is generating topic trends from the results of topic modeling. For each topic, we calculated a trend over the broadcasting days and matched with corresponding viewer ratings. The last step is time-series analysis with VAR model for finding the relationship between the topics and the viewer ratings.

TOPIC TREND ANALYSIS OF SOCIAL MEDIA ON DRAMAS

Crawling Documents from Blogs

The broadcasting days of the most dramas are irregular as mentioned before while there are social media conversations every days of the week. Therefore, we crawled Blog documents written at the day after broadcasting for future use in addition to the days of broadcasting. For example, we crawl the documents of Monday, Tuesday, and Wednesday if the broadcasting days are Monday and Tuesday. It is interesting that the number of documents is not always proportional to viewer ratings. For example, "Wang Family" has the highest viewer rating but the lowest number of documents. On the other hand, "Misaeng" which means incomplete life has the lowest viewer rating while it has quite many documents compared to other dramas.

Table 2. Results of Crawling from Blogs

Title	# of Broadcasting	# of Viewer Rating	# of Documents	Avg. # of Documents per Day
Wang Family	50	75	4,818	64.2
Empress Qi	51	77	10,130	131.6
Dr. Stranger	20	30	5,249	175.0
My Love from the Star	21	32	43,913	1372.3
Misaeng	20	30	15,787	526.2

Topic Modeling Results

A topic model is a generative model for documents [13]. One of assumptions of topic modeling is that documents are mixture of topics and a topic is a probability distribution over words. That is, different documents can be generated by picking a set of topics with weights given to each topic. Therefore, we can infer topics that were most likely responsible for producing a set of documents. One of issues in our research was determining the number of topics. It is hard to interpret the topics if there are too many topics, and the topics are very broad if there are too few topics [13]. Mallet shows topic number, overall weight and words for each topic as results in addition to document name, topics and weights for each document. Therefore, we chose 5 out of 10 topics generated from the results based on the overall weight of each topic. That is, we removed the topics with lower weight. After that, we named topics but this process is somewhat subjective. Table 3 shows a topic modeling result of "My Love from the Star". It is interesting that we could identify clear topics that show different subjects as shown in the table. The topics "Food" and "Fashion" are very clear. Songee, Dominjun, Semi and Jaekyong are main characters of the drama, and the remaining words of "Story/Roles" are about the story of the drama. For "Drama – General talk", "The heirs" and "The thieves" are titles of other drama and movie, and other words are general terms about drama. The words of "Daily Life" are not about the drama. We could extract seven common types of topics from all target dramas.

Table 3. A Topic Modeling Result of "My Love from the Star"

Topic Title	Weight	Words of Topic
Food	0.13184	Food, price, cheese, strawberry, pizza, ice coffee, sweet potato, cafe …
Fashion	0.12519	Overcoat, price, style, sunglasses, dress, jacket, fashion, skirt, shirt, …
Story/Roles	0.14532	Songee, Dominjun, Semi, Jaekyong, alien, star, epilogue, kiss, supernatural power, …
Drama – General talk	0.18174	Drama, viewer rating, casting, review, actress, The heirs, The thieves, acting ability …
Daily Life	0.20568	You, today, mobile, thought, photo, love, We, human, sister, time, one day, friend, now…

Topic Trends of Social Media and Viewer Ratings

Topic trends provide changing trends of topics over time [9]. First, we calculated the average of weights of each topic over the documents of the same day. After that, we multiplied the average with the number of documents to reflect the effect of the intensity of social buzz. As the results of topic trend analysis, we could get the popularities of each topic for each broadcasting day for every drama. Table 4 shows the time-series vectors that consists of broadcasting date, viewer ratings and popularities of topics. In the table, each topic including viewer ratings has time series because those varies over broadcasting days.

Table 4. Trend of Topics and Viewer Ratings of "My Love from the Star"

Broadcasting Date	Viewer Ratings	Popularity of Each Topic				
		Daily Life	Story/Roles	Fashion	Food	General
2013-12-18	15.6	73.5	27.8	26.2	21.0	173.5
2013-12-19	18.3	176.0	93.0	134.4	45.2	465.4
2013-12-25	19.4	78.4	30.4	30.4	21.5	76.9
2013-12-26	20.1	168.1	121.6	139.8	63.9	213.7
2014-01-01	22.3	92.5	30.4	32.8	27.5	68.0
2014-01-02	24.6	208.9	125.8	162.0	72.6	165.0
2014-01-08	24.1	220.6	112.8	126.1	80.2	189.8
2014-01-09	24.4	351.8	257.2	223.4	138.7	263.5
2014-01-15	23.1	298.5	123.5	159.6	128.4	225.7
2014-01-16	24.4	415.9	357.9	260.2	178.5	310.3
2014-01-22	24.5	406.7	218.2	239.8	170.1	245.5
2014-01-23	26.4	646.5	502.8	557.1	239.1	420.4
2014-01-29	24.8	332.2	136.1	205.9	109.2	281.8
2014-02-05	25.7	396.5	171.1	215.5	161.1	222.5
2014-02-06	25.9	533.9	340.7	326.2	209.1	280.0
2014-02-12	25.7	480.1	229.3	203.9	183.1	284.5
2014-02-13	27	534.5	323.4	327.5	224.9	343.7
2014-02-19	27.4	449.9	221.1	207.3	164.2	263.8
2014-02-20	26.7	551.0	345.2	299.2	215.4	296.3
2014-02-26	26	436.2	190.5	191.9	166.9	354.5
2014-02-27	28.1	825.4	501.5	345.8	288.7	575.4

Figure 2 shows the topic trends and the change of viewer ratings of "My Love from the Star". There are three outstanding peaks of topic weights in the graph because of suddenly increased number of documents of the day. Except for the peaks, the relationship between topics and viewer ratings is easily detected in the graph. It will be interesting if we look through topics of the peaks in detail to find out causes for our future research.
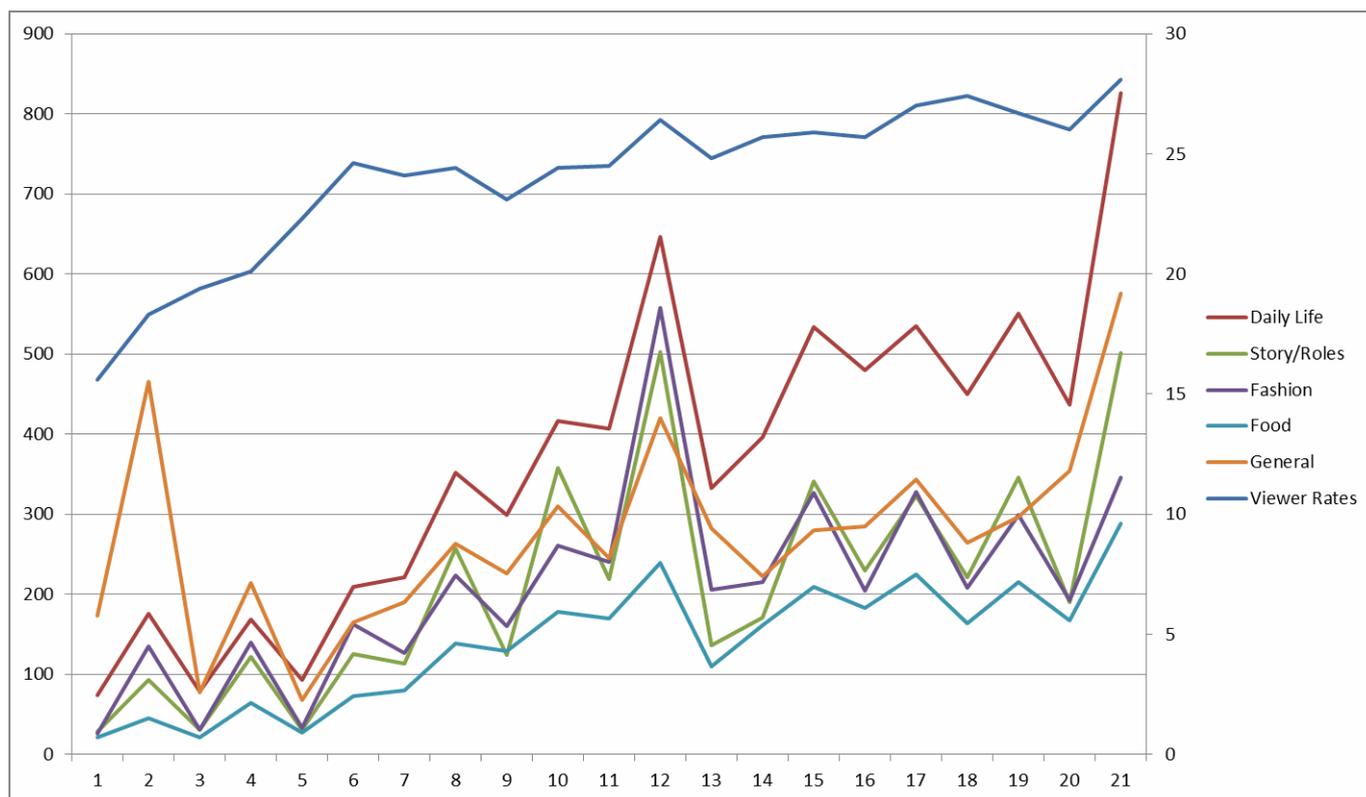
Figure 2. Graph for Table 4

VAR ANALYSIS FOR TOPC TRENDS AND VIEWER RATINGS

Issues of Using VAR

We used unstructured model for our basic structure in VAR model because of the irregular days of broadcasting. Adjusted topic weights and the viewer rating of each day were assigned for endogenous variables. Topic weight is a relative value among topics of a document. Therefore, we multiplied the average topic weights with the number of documents to reflect the quantity of social media buzz as we mentioned before. As the result, we could get better results in Granger causality tests and relationships among topics. The most important issue in our analysis is the insufficient observations. Therefore, we limited the number of topics of each drama for three or five. Moreover, we estimated viewer ratings of the day after broadcasting. For example, we calculated the viewer rating of Wednesday which was not a broadcasting day with the average of viewer ratings of Tuesday and the next Monday.

VAR Analysis Results

Table t shows the VAR results of "My Love from the Star". The topic "Story/Roles" negatively affects other topics and viewer rating of the next period. The more topics about roles and stories (that are specific to drama) are, the less other topics are in most cases. The trend scores are relative, so it means that social media buzz tends to focus on specific contents of a drama as time goes by. On the other hand, viewer rating positively affects every topics and viewer rating itself as we expected.

There are some common observations on all dramas. First, in most cases, viewer rating positively affects topics. Second, Topics hardly affect viewer rating. We guess it is because Blog is a kind of diary, so it is hard to influence other people. Third, the topic "Story/Roles" negatively affects other topics and viewer ratings in most cases. Fourth, the topics such as background and philosophy seem to stimulate other topics.

Table 5. VAR Results of "My Love from the Star"

	STORY_ROLES	DAILY_LIFE	FOOD	VIEWER_RATING
STORY_ROLES (t-1)	-0.811	-0.874	-0.780	-0.009
	[-2.404] *	[-2.855] *	[-2.990] *	[-3.175] *
DAILY_LIFE (t-1)	1.456	1.451	1.114	0.001
	[1.774]	[1.946]	[1.754]	[0.225]
FOOD (t-1)	-0.640	-0.658	-0.367	0.005
	[-0.760]	[-0.860]	[-0.563]	[0.674]
VIEWER_RATING (t-1)	17.476	23.372	21.297	0.890
	[2.103] *	[3.095] *	[3.312] *	[13.295] *
R-squared	0.470	0.580	0.619	0.939

CONCLUSION

We developed a method to analyze the relationship between social media activities and TV viewer ratings. In detail, we conducted topic model analysis on 5 popular Korean famous dramas, topic trend analysis on the dramas and VAR analysis on topics and viewer ratings. We found some meaningful relationships between the topics discussed in social media and the viewer ratings. For example, viewer rating positively affects topics and topics hardly affect viewer rating. Moreover, the topic "Story/Roles" tends to negatively affect other topics and viewer ratings in most cases.

There is a lot of future research topics. First, we need to control irregular broadcasting days even though we used some techniques in the study. This will the most important issue of our next research. Second, we need to increase the degrees of freedom in order to analyze more topics. Third, we need to add control variables such as viewer ratings of TV News to enhance the model. Last, we need more in-depth analysis on topics to get better interpretation of topics.

ACKNOWLEDGEMENT

REFERENCES

[1] Carr, D. (2013) 'Giving viewers what they want, [Online]', *The New York Times*, available at http://www.nytimes.com/2013/02/25/business/media/for-house-of-cards-using-big-data-to-guarantee-its-popularity.html?_r=0 (Accessed 15 May 2015).

[2] Cesar, P. & Geerts, D. (2011) 'Understanding social TV: A survey', *Proceedings of the Networked and Electronic Media Summit (NEM Summit 2011),* Torino, Italy, pp. 94-99.

[3] Cesar, P. & Geerts, D. (2011) 'Past, present, and future of social TV: A categorization', *Consumer Communications and Networking Conference (CCNC)*, IEEE, 2011. IEEE 2011, pp. 347-351.

[4] Deb, R. (2012) *Social TV,* [Online], TVB., available at http://www.tvb.ca/pages/deb+roy+tv+day+2012_htm (Accessed 15 June 2015).

[5] Jung, H. (2014) 'Economic ripple effect of "My love from the star",[Online]', *Cultural Tourism,* available at https://www.kcti.re.kr/webzine2/webzineView.action?issue_count=32&menu_seq=4&board_seq=1 (Accessed 13 June 2015).

[6] Kushin, M. J. & Yamamoto, M. (2010) 'Did social media really matter? College students', use of online media and political decision making in the 2008 election', *Mass Communication and Society,* Vol. 13, pp. 608-630.

[7] Lee, H. (2014) '2014 Drama Rank, [Online]', *NEWSPIM*, available at http://www.newspim.com/view.jsp?newsId=20141222000007 (Accessed 21 May 2015).

[8] Lee, S., Baker, J., Song, J. & Wetherbe, J. C., (2010) 'An empirical comparison of four text mining methods', *System Sciences (HICSS), 2010 43rd Hawaii International,* Conference on, 5-8 Jan. 2010, pp. 1-10.

[9] Matsubara, Y., Sakurai, Y., Faloutsos, C., Iwata, T. & Yoshikawa, M. 'Fast mining and forecasting of complex time-stamped events', *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 2012*, ACM, 271-279.

[10] Michaelidou, N., Siamagka, N. T. & Christodoulides, G. (2011) 'Usage, barriers and measurement of social media marketing: An exploratory investigation of small and medium B2B brands', *Industrial Marketing Management,* Vol. 40, pp. 1153-1159.

[11] Roettgers, J. (2014) *Let's face it: social TV is dead,* [Online], available at https://gigaom.com/2014/01/29/lets-face-it-social-tv-is-dead.

[12] Sims, C. A. (1980) 'Macroeconomics and reality', *Econometrica: Journal of the Econometric Society*, pp. 1-48.

[13] Steyvers, M. & Griffiths, T. (2007) 'Probabilistic topic models', *Handbook of latent semantic analysis,* Vol. 427, pp. 424-440.

[14] Team, D. M. (2014) '2014 SNS Usage Report, [Online]', *DMC Media,* available at http://www.slideshare.net/BeckyKim/2014-sns-37219822 (Accessed 17 April 2015).