The Prediction of Sales Volume and WoM Effect through Network Structure and Topic Modeling on Social Media

Completed Research Paper

Bongjin Sohn

Hyuna Lim

Jaewon Choi

Abstract

Social media has emerged as a diverse platform and its influence has gradually improved. The company recently carried out a strategy to build a network of communication between users through social media. In this study, we analyze the empirical analysis of social media communication structure and predicting the product sales volume based on the literature review of the existing media theory. Social media can be divided into two dimensions of the communication process (Conveyance vs. Convergence). The Conveyance means information transmission, and that purpose is to convey information broadly. The Convergence means gathering opinion and aiming to enhance knowledge by sharing of users. Using social media, users are gained by social capital through the structural, cognitive and relational structure. As a result of the study, we drawn three communication structure affecting the company’s sales volume. In addition, the factors can predict the sales volume.

Keywords: Social media, Media Synchronicity Theory, Conveyance, Convergence, Data mining

Introduction

The diffusion of social media has improved on the accessibility of the information among users for the spread of internet and mobile environment. YouTube has become growing the platform can be connected by all over the world users with a mobile device that has provided various contents and languages. The company’s values are assumed to be about 10 trillion after Google Inc. M&A (Adamopoulos et al. 2018).

Social media can be made available for contact of output performance paying lower cost and more profit to users. The consumer can be able to use other users’ making contents text, picture, and video themselves. By using the others making production, the consumers can be easy to get WoM effect such as decision making and acquiring valid information (Miranda et al. 2015).

One of the best examples of a company is Tesla Inc. The company has been called the innovation of automotive industry. Because the incorporation preoccupied innovative category on electronic motor
vehicle industry overcoming the Chasm is difficult to get over the high-tech industries. The factor of success in the marketplace is operating experience marketing. Tesla Inc. gave the ‘Hedonic’ impression, their production as against the most of high-tech firms gave the ‘Utility’ impression. Tesla Inc. has a commitment to provide a real user experience on their offline booth and encourage to share social media for other social community users (Adamopoulos et al. 2018; Miranda et al. 2015; Gerald et al. 2014; Dewan and Ramaprasad 2014).

Therefore, the power of social media is the most influential channel between the company and social media user. They must make the best use of social media for getting the necessary information, sharing, diffusion. For this reason, this study focuses on the effect of communication on social media by analyzing social network structure, text mining and prediction analysis method (Nahapiet and Ghoshal 1998).

The antecedent studies of media theory have been developed such as Media Richness Theory and Media Synchronicity Theory. However, antecedent studies are mostly verification using limited methods such as exploring research design or quantitative research by questionnaires design (Dennis et al. 2008). Therefore, we need to apply extensive diverse methodology utilizing leaving users’ traces such as text, social relationship. This study tries to verify the value of unstructured data based on social media using social network analysis, text mining, artificial neural network and statistical method (Andrade and Doolin 2016).

The research questions of this study are as follows: First, how the difference of communication effect between expert community (i.e. Wikipedia) and non-expert community (i.e. YouTube) based on social media? Second, how the difference of social capital through communication process between Conveyance(non-expert) and Convergence(expert) based on Media Synchronicity Theory? Third, how does the network structure consist of social capital on social media affect the prediction of the sales quantity of production?

**Literature Review**

**Social media**

Social media has been used since the 2000s when Web 2.0 was developed, and now it has become a channel that most used channel. The term of the social media was first used by arguing Chris Shipley in 2004 and academic research on the terms of social media has continued to date (Miranda et al. 2015; Wang et al. 2016).

Social media does not follow the mechanism of production and consumption of existing products and has the characteristic of growing itself through participation among users because it changes organically. Contrary to the one-sided nature of traditional communication media, the user forms an interactive multimedia structure that sometimes becomes a producer and sometimes a consumer (George et al. 2018).

Social media definitions are defined as "the space where users can express who they are through participation in an online platform environment, create their own content, acquire and spread the information through sharing interactive information between individuals." (Dewan and Ramaprasad 2014)

Social media classification is largely divided into four types and are communication, entertainment, collaboration, content sharing, and communication (Gerald et al. 2014). The antecedent research related to social media is mainly based on research on the motivation for social media use (Shore et al. 2018).
The Media Theory

Media Synchronicity Theory

The Media Richness Theory (MRT) has achieved theoretical integration and expansion through existing related media theories. There is a difference in the possibility of information transmission, amount of information, and kind of information depending on the type of media channel. Thus, there is a difference in the richness provided between the media, and it is reasonable to choose the media that suits the situation for effective communication (Daft and Lengel 1986; Dennis and Kenny 1998).

To overcome the theoretical limitations of the richness theory (MRT), the Media Synchronicity Theory (MST), which focuses on media attributes and capabilities, has been argued. MST has been argued focusing on how the users of information and the recipients of information can acquire information concurrently (Dennis et al. 2008). Information processing ability, and synchronicity among users influence the effect of communication and collaboration. (Burke and Cildamabram 1999; Addas and Pinsonneault 2018).

Synchronicity claims that users can often feel asynchronous even if they interact with the same time zone through the media (Mirdanda and Saunder 2003). Therefore, there may be a time difference depending on the time recognized by the individual, and there is synchronicity. A comprehensive definition of synchronicity can be assessed as the degree of competency of the media that controls communication (Dennis et al. 2008).

Synchronicity has demonstrated that it has superior explanation compared with the media richness theory. MST has suggested factors affecting five communication outcomes to understand the media properly (Dennis et al. 2008). The Transmission Velocity is an element that increases the synchronicity of the media, and it determines the synchronicity through immediate feedback on the message delivered by the sender. Parallelism is defined as sharing the information of the sender through the media to various users, and the synchronicity decreases when the parallelism increases by being shared with many users. Symbol Set includes writing, voice, and images as elements of encoding, decoding and symbolizing information to grasp various situations and communicate information (George et al. 2013).

Rehearsability refers to the ability of the sender to modify and confirm the information by confirming that the information is well processed. Reprocessibility refers to the process of decoding information when the recipient receives and receives information, and can be reinterpreted different from the intention sent by the sender according to the interpretation of the prisoner.

On media synchronicity theory, the Communication Process is divided into two dimensions: Conveyance (information delivery) and Convergence (opinion gathering). Conveyance is a process of delivering a variety of new information, and Convergence is defined as a process of collecting and sharing opinions (Dennis et al. 2008).

Although media-related studies based on media synchronicity theory are ongoing, most of the researches in the developmental aspect of the theory exist, and the research that is empirical analysis by applying the theory is still insufficient (Zhang et al. 2016; Minas et al. 2014). Therefore, this study aims to measure the communication effect in actual social media through the theoretical expansion of media synchronicity theory and to check whether it has a relationship with performance.
Communication Process based on MST

The situational requirements of users who use the media are also presented in a wide variety of ways, as the types and types of media vary depending on the characteristics required by each. Therefore, depending on how the media is used in any situation, the effect and performance of communication may vary considerably. In the previous study, we described the communication situational requirements in the context of group decision making in two major processes: Conveyance (information delivery) and Convergence (information convergence) (Robert and Dennis 2005).

The information delivery process (Conveyance) is a communication process aimed at sharing and delivering a lot of various forms of information through the media. Since the information delivery process simply shares information among users, the community requires high quality information and features information diffusion in various ways. From the point of view of the processing perspective, conveyance shows a relatively slow transfer rate. In the process of transmitting information, the sender is a process that requires relatively low synchronicity because it simply transmits information through the media and does not think about feedback or response afterward (Brown et al. 2010).

The opinion gathering process (Convergence) is a communication process aimed at collecting information from each user through the media and obtaining consensus through negotiations between information providers and prisoners based on the collected information. The process of collecting opinions refers to the process of collecting opinions and developing new knowledge by collecting pieces of prior experience and knowledge that participating users have. Therefore, the rate of feedback between users is relatively fast, and the opinions of the other party are achieved through frequent communication of 1 to 1 within the group. Therefore, it requires relatively high synchronicity and is a process of immediate response and modification (Yadav and Pavlou 2014).

Depending on the communication process, knowledge sharing on the media occurs in different ways, but the two-dimensional communication process is not an exclusive element (Dennis et al. 2008). So, it can be Conveyance Process, and it can be Convergence Process, depending on the situational factors (Jiang et al. 2010).

The online community can be classified as a professional and non-professional knowledge community based on the information provided. Professional communities share norms and values among users who make up communities and improve their knowledge through critical thinking (Scribner et al. 1999). On the other hand, in the case of non-professional communities, users who use online can voluntarily handle various information and treat it free of charge (Boughzala et al. 2012).

The communication process derived from the theory of synchronicity varies depending on the difference between professional and non-professional communities. Professional community refers to a community with a process of sharing and developing knowledge with a common purpose through theoretical consideration and has the characteristics of the Convergence Process. non-professional communities refer to a community that has a process of easily acquiring, sharing and spreading information, and has the characteristics of a Conveyance Process.

Communication Process and Social Capital

Knowledge sharing has the power to preoccupy or exercise as a power, according to the amount and quality of knowledge owned by users who use the community. So, you get social capital through contact with the user who owns the knowledge (Adamopoulos et al. 2018). On social network sites, users are productively engaged in creating and exchanging intellectual capital, such as knowledge, through interpersonal interactions (Wasko and Faraj 2005). Social capital is defined as “a total of actual and potentially included resources, which can be acquired through a network of relationships owned by individuals or groups, and derived from them” (Mojdh et al. 2018).

Social capital is divided into three dimensions: Structural, Cognitive, and Relational (Nahapiet and Ghost 1998). Structural dimensions mean the type of connection between users in the network and the pattern of connection between users (Nahapiet and Ghost 1998). The Structural dimension can be
measured by the type of connection between members in a group and can be measured by the Tie Strength or Structural Hole of the entire group’s network (Nahapiet and Ghost 1998).

The Cognitive dimension is defined as an interpretation and expression of the resources formed between users in the group, and a common understanding and perspective of all members of the network. Therefore, Clique and Network Modularity are available (Gerald et al. 2014). The relational dimension is defined as ‘the degree to which assets are created and assets are utilized through the quality and relational characteristics of the network among members in the group’ (Nahapiet and Ghost 1998). The relational dimension is a factor that can identify reciprocity and interaction of connection relationships and reduce social distance costs, and can identify network forms and identify relationships among members of the group through key indicators of network centrality (e.g. Degree centrality, Betweenness centrality, Closeness centrality) (Gerald et al. 2014).

It is necessary to check the quality of the network in the group through three dimensions of social capital and to confirm specifically how it affects the communication effect. There is an antecedent study on capital acquisition according to the level of social capital in social media, but research is insufficient (Gerald et al. 2014). The social capital that is based on media theory and formed according to two dimensions of the communication process will be different (Conveyance vs. Convergence). In this study, we tried to distinguish two dimensions and to analyze three dimensions of social capital according to each process level through social network analysis (Davison et al. 2014).

Social capital is also acquired differently depending on the two dimensions of the communication process. Therefore, this study is to measure the three dimensions of social capital by dividing them into two dimensions of the communication process (Conveyance vs. Convergence) (Borgatti et al. 2013). First, Network Density, which is a variable that explains the structural dimension, extends the information to more people more widely, so that it can construct a relatively low density and structure the Conveyance Process. In the case of Structural Hole, it is an extended concept of weak strength between nodes, and it is an indicator suitable for Convergence Process because it can measure the structural role of intermediating a network with opinions in each field in the professional knowledge community (Ludwig et al. 2014).

Second, Clique, a variable that describes the Cognitive dimension, is suitable for conveyance, where rapid information diffusion is important because they form a group with cognitive synchronicity and perform functions to speed up information diffusion. In the case of Network Modularity, information is shared firstly with nearby neighbors in spreading information within the network, so the higher the Modularity, the higher the proportion of subgroups and the faster the information spread (Zhang and Venkatesh 2013).

Third, the Network Centrality can explain the Relational dimension of social capital through the connection of users in the social media network. Degree Centrality can explain the spread of information because it shares resources and, in the case of high centrality, it makes the flow of resources desired. Closeness Centrality requires inter-node information sharing and negotiation coordination for knowledge development, so for expert knowledge, independence is shown in the process of feedback. Betweenness Centrality is suitable for convergence because it requires an interventional role in the process of connecting and developing nodes (Borgatti et al. 2013).

The two dimensions of the communication processes (Conveyance and Convergence) are as follows. Non-specialized community (Conveyance) was measured by Network Density, Clique, Network Modularity, and Degree Centrality. The professional community (Convergence) was measured by Structural Hole, Closeness Centrality, and Betweenness Centrality.

Research Methodology

Research Procedure

This study aims to divide the communication process of social media into two dimensions (Conveyance vs. Convergence) by establishing a theoretical basis for media theory (MST) and to apply the
characteristics to professional and non-professional communities to check the WoM effects of the community through network structural and relational analysis and to check the causal relationship between the values and the company's outcomes.

To empirical analysis the study, a total of six steps were conducted. We collected online unstructured data. Based on refined data, we conducted social network analysis to derive communication structural indicators. Through Topic modeling algorithm, we derived potential topics that are mainly discussed in social media, confirming meaning, and showing trends over time.

Based on the derived network indicators, statistical analysis was conducted to prove how the WoM effects of social media affect outcomes and whether the communication structural indicators can be identified as factors affecting performance. In addition, we implemented modeling to predict outcomes through artificial neural network analysis.

![Figure 1. Research procedure](image)

**Data Collection**

For analyzing this study, we selected Conveyance communication processes and Convergence communication processes on social media. In the case of Conveyance processes, we selected YouTube, which is the most widely used, easy to share and transmit information. In the case of Convergence, we selected Wikipedia, which is social media with expertise based on collective intelligence.

In the two social media channels (i.e. YouTube, Wikipedia), we compared the automotive electronics corporation of Tesla and Nissan. On YouTube, we collected users and text of the title, content, and comments of all posts including 'Tesla Model S' and 'Nissan Leaf'. The period is from the first quarter of 2014 to the third quarter of 2018, collecting data for a total of 19 quarters. Wikipedia collected data from the revisions network edge list of the historical view on the Wiki page of 'Tesla Model S' and 'Nissan Leaf'. The period also is from the first quarter of 2014 to the third quarter of 2018, collecting data for a total of 19 quarters.

Tesla Model S can be seen that it responds to the issue very sensitively. The reason for comparing the two automotive electronics have a lot of popularity and steady sales volume and rank in the top rankings in reports such as the electric car market report. Conveyance (YouTube) collected the title, content, and users of a total of 12,463 videos and collected 738,005 comment data. Convergence (Wikipedia) collected the relational data of a total of 3,555 revisions.
Table 1. Data collection on social media

(Period from the 1Q of 2014 to 3Q of 2018)

<table>
<thead>
<tr>
<th>Index</th>
<th>Conveyance (YouTube)</th>
<th>Convergence (Wikipedia)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tesla Model S</strong></td>
<td>Videos 7,218 (585,979 comments)</td>
<td>2,772 revisions</td>
</tr>
<tr>
<td><strong>Nissan Leaf</strong></td>
<td>Videos 5,245 (152,026 comments)</td>
<td>1,183 revisions</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>Videos 12,463 (738,005 comments)</td>
<td>3,555 revisions</td>
</tr>
</tbody>
</table>

We used secondary data, which is an indicator of the actual performance of electric vehicle companies, for analysis using the data presented in official reports (Push EVs.com).

Figure 2. Secondary data of automotive electronics outcomes (Tesla Model S vs. Nissan Leaf)

Data Analysis and Results

Social Network Analysis

We modified the collected data to a unit that can analyze the social network through preprocessing, and changed the direction graph to a non-directional graph (symmetric) and used it for analysis. The data collected from Conveyance (YouTube) and Convergence (Wikipedia), two dimensions of the communication process, were used to analyze through a total of six network structure indicators based on three dimensions of social capital. The purpose of this study is not to derive through the microscopic environment but to examine the utility of the structural indicators at the macroscopic level by using the variables of the media channel itself. Six network indicators including structural voids and modularity were used to extract indicators that could explain the results against the secondary data.
### Table 2. Operation definition

<table>
<thead>
<tr>
<th>Index</th>
<th>Communication Process</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural dimension</td>
<td>Structural Hole; SH</td>
<td>Structural Hole are empty spaces between unconnected nodes, which are important in the network process of collecting various opinions among members who have new ideas.</td>
</tr>
<tr>
<td>Cognitive dimension</td>
<td>Clique</td>
<td>Clique is a completed graph with three or more nodes connected completely, and in Conveyance, the completer the graph, the more common subgroups are created, and the information is delivered faster.</td>
</tr>
<tr>
<td></td>
<td>Modularity</td>
<td>Modularity is the ratio of the actual subgroup connection to the randomly connected expectation, and the higher the modularity, the greater the probability of subgroup formation, and the faster the information transfer through common interests.</td>
</tr>
<tr>
<td>Relational dimension</td>
<td>Degree Centrality; DC</td>
<td>Conveyance</td>
</tr>
<tr>
<td></td>
<td>Closeness Centrality; CC</td>
<td>A network with high closeness centrality is a network with prominent independence between nodes, and it is necessary to negotiate and coordinate each node. Therefore, the process of negotiation and coordination is essential to gather opinions.</td>
</tr>
<tr>
<td></td>
<td>Betweenness Centrality; BC</td>
<td>In the case of betweenness centrality, there are the nodes, which plays a role in mediating the connection, and is a necessary network indicator for mediating the flow of contents in the process of gathering opinions relatively.</td>
</tr>
</tbody>
</table>

### Text Mining

Using the Latent Dirichlet Allocation (LDA) algorithm, we analyzed text data based on YouTube. A total of five topics was specified to summarize the main topics of the 'Tesla Model S' and 'Nissan Leaf' documents. The top 30 keywords have been specified to specify the topic of the document. In Tesla Model S, five topics were derived: Automatic performance, User experience, Automatic openness principle, and Battery performance. Among the keywords from the Tesla model S included many meaning related to the user experience are derived. Tesla's marketing strategy is to use experiential marketing in social media, so it seems that many keywords related to user experience are derived.

Nissan Leaf were derived by major topics such as Automotive performance, Consumer response, Battery pack, Automotive operation principal, and Charging station. Unlike Tesla Model S topics, Nissan Leaf has a lot of performance keywords such as major batteries and charging stations, and fewer keywords related to social media characteristics such as user experience are derived. Since Nissan is a relatively low-priced electric car, it is assumed that more functional performance keywords are derived because consumers tend to buy and consider attributes.

Based on our extracted topics, Dynamic Topic Modeling (DTM) was conducted to identify the traffic patterns of 19 quarters. Dynamic topic modeling allows to identify which topics are important quarterly. In addition, by comparison with the sales volume of the electric car, which is the secondary data, we tried to identify the main topics related to the sales quantity by extracting the rising topic when the sales volume increased.
By summarizing the results, Tesla can improve the user experience through experiential marketing and confirm that consumers respond through comparison of car performance with other companies. Nissan's sales volume is related to the WoM effects of charging stations and automobile performance, and the topics that consider the automotive specifications and infrastructure of Tesla are more practical.

**Linear regression**

Through the extracted communication network indicators, statistical approaches were conducted to confirm whether structural and relational indicators are causal to company's outcomes. For two
dimensions of the communication process (Conveyance & Convergence) as a prior analysis, we analyzed t-test of two dimensions in groups to determine whether the network analysis results indexes distinguish dimensions well (i.e. Conveyance and Convergence). As a result of the analysis using t-test, all six social capital factors were found to have average differences between communication processes. Therefore, according to the difference in the communication process, there is a structural and relational difference of the network and the WoM effect is also different.

A linear regression analysis was carried out to determine whether a total of six indicators affects the sales quantity. Since the dependent variable is a discrete variable, the variable is changed to a continuous variable considering the heteroscedasticity of the variable and analyzed. In other words, for linear regression, we analyzed to take the natural log value of sales volume to the dependent variable (SalesVolume).

### Table 3. The result of linear regression

<table>
<thead>
<tr>
<th>Index</th>
<th>β (Standardized)</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CY_Modularity</td>
<td>0.573</td>
<td>2.184 **</td>
<td>0.037</td>
</tr>
<tr>
<td>CG_SH</td>
<td>0.521</td>
<td>2.358 **</td>
<td>0.025</td>
</tr>
<tr>
<td>CG_CC</td>
<td>0.614</td>
<td>2.266 **</td>
<td>0.031</td>
</tr>
<tr>
<td>CG_BC</td>
<td>-0.184</td>
<td>-1.015</td>
<td>0.318</td>
</tr>
<tr>
<td>CY_Clique</td>
<td>0.368</td>
<td>1.715</td>
<td>0.097</td>
</tr>
<tr>
<td>CY_DC</td>
<td>0.426</td>
<td>1.588</td>
<td>0.123</td>
</tr>
</tbody>
</table>

Dependent variable: Ln(SalesVolume)
Adjusted R Square: 0.223
VIF: 1.486~3.304 (<10)
Sig. 0.05>p *, 0.01>p **

The result of analysis showed that the network indicators affecting sales volume were Conveyance_Modularity (CY_Modularity), Convergence_Structural Hole (CG_SH), and Convergence_Closeness Centerality (CG_CC).

### Artificial Neural Network Analysis

We checked whether the three indicators derived from regression analysis can predict sales volume, we implemented artificial neural network (ANN) analysis modeling, one of the representative methods of forecasting analysis. To confirm the predictive power of three statistically significant indicators, we implemented six artificial neural network models (from 6 to 1 communication structures based on linear regression result).

For analyzing ANN, the multilayer perceptron was modeled by combining the values of Tesla’s 19 quarters and Nissan’s 19 quarters. The data to be trained in the model consists of data that learns 70% of the total data and predicts the remaining 30%.

The verification of the prediction modeling was verified through the coefficient of determination and the RMSE (Root of the Mean Square Error) that measured the error rate. The closer the coefficient of determination is to 1, the better the prediction model. The closer the RMSE is to zero, the better the prediction model. As a result of the analysis, it was confirmed that the error rate was lowered while the determinant coefficient was high when the 3 variables derived from the statistical test were high.

### Table 4. The result of Artificial Neural Network Analysis

<table>
<thead>
<tr>
<th>Index</th>
<th>Coefficient of determination</th>
<th>RMSE</th>
</tr>
</thead>
</table>
Conclusion

The study was based on media theory to identify the difference of communication effect between the Conveyance (information delivery) process, and Convergence (gathering opinion). Conveyance is a process that delivers information to many users for information delivery and requires a large network and wide scalability. Therefore, it is aimed at spreading the information acquired by the non-professional community to as many others as possible and spreading it widely.

Convergence, on the other hand, refers to the process of collecting information that each user has and creating new advanced knowledge for collecting opinions. Therefore, because it forms advanced knowledge, it is divided into community expertise, and it is aimed at forming meaningful knowledge by negotiating, coordinating and mediating opinions belonging to the professional community.

This study has several academic implications through theoretical expansion by reviewing and utilizing existing prior studies. First, we tried to expand the theory with unstructured data based on the existing media theory (i.e. MRT, MST). Second, we expanded the communication process, which is a situational requirement of MST, social media applied to explain community expertise and explain how knowledge sharing activities are done. Third, we analyzed factors and identified causal relationships through analysis of network structures that were not well utilized in existing consumer behavior and marketing research. Finally, we expanded the scope of research on predicting outcomes indicators by conducting neural network prediction analysis based on communication structural factors.

This study has some practical implications through empirical analysis. First, unlike traditional marketing, social media can be used to respond to the customer’s responses in advance. Second, it is necessary to continuously induce WoM in the enterprise for the development of powerful opinion of leaders and collective thinking in the market, rather than using basic network analysis indicators based on media theory. Third, it can be used as a predictor's tool to identify consumer response immediately in social media through structural analysis and forecasting analysis of social networks. Finally, it is necessary for the company to consider the 'Bait strategy' to improve the quality of the network itself by participating in the network directly, not the marketing that is being fed.

The limitations of this study are as follows. To analyze the two dimensions of Conveyance and Convergence, we limited social media to YouTube and Wikipedia. Second, it was not possible to adjust the control variables in addition to the network structural factors of the community. Third, to measure the effectiveness of communication of new electric car industry, the two companies and vehicles were selected and analyzed.

Based on this study, we can carry out further research through several expansions. First, in addition to electric cars, which are one of the new industries, we can expand our research on how to overcome strategic chasms by identifying the structure of communication. Second, it can be applied to various fields related to social media such as fandom, subculture and new product launches as well as innovative products or new industries, and theoretical expansion can be brought about.

References