Big data Analysis on Gambling related News in South Korea

HyeJung Moon
Ph D. of Public Policy
CLI, ILP, SeoulTech Unv.
Seoul, South Korea
hyejung.moon@gmail.com

SungKyung Kim
Master of Public Policy
Crossent
Seoul, South Korea
gbible45@gmail.com

Thomas Yang
Bachelor of Arts
Intralot S.A.
Seoul, South Korea
thomas.yang@intralot.co.kr

KwanWoo Kim
Bachelor of Computer Science
Cooperate Lottery Institute
Seoul, South Korea
kyle_kwkim@naver.com

Abstract— **The purpose of this study is to understand the recognition of gambling industry through the semantic analysis of news data on lottery, sports betting, horse racing and casino that was reported between 1990 to 2015 in South Korea. This paper revealed the difference between journalists' intention and public's perception about news by analyzing the frequency and connectivity of news with framing and public's interest through semantic network analysis and explored the policy characteristics and innovation task. The result of analysis, news on lottery game mainly has been reported social issue related with win such as 'winning number', 'prize money', 'suspicion of manipulation' and etc. News on sports betting has been reported mandatory information related with business project and illegal site such as 'bidding', 'illegal site', 'sales target' and etc. News about horse racing has been reported the information about the business advertisement such as 'online race track' and 'promotion'. Lastly, casino related news has been reported 'major information' such as illegality', 'gambling place' and 'foreigner'. As a result of times series analysis, news about casino in the 1990s, news about lottery in the 2000s and news about horse racing in 2010s have been increased. Public's interest also has been moved to 'business scandal', 'winning game', 'citizens´ campaign' and etc. Gambling related news has been classified by four types, 1. advertising publicity(horse racing), 2. mandatory information(sports betting), 3. social issue(public agenda, lottery), 4. major information(casino). We could get the insight that news can be formed a public agenda, when news is reported as a social issue with high frequency and public's interest like lottery related news.**

Keywords-gambling; lottery; news; public policy; big data analysis; frame theory

I. INTRODUCTION

According to the National Gambling Control Commission, the sales volume of gambling industry in South Korea has been almost doubled to 20.5 billion USD(When we calculate that the exchange rate is 1 US$ equals 1,000 KRW.) in 2016 compare to 12.5 billion USD in 2006 [7]. It shows that the Koreans have very strong interest towards gambling. In the case of Lotto, the sales increased 10 times since the first issuance on December, 2002. News from major media also has shown high interest. The news coverage rate of Mediagaon from 1990 to 2015 on lotto, sports betting, and casino which are the major gambling industry showed high interest. The Mediagaon(The name of Mediagaon was changed to BigKinds(www.bigkinds.or.kr) at April 2016.) is the company which has the news database of 61 media companies such as national daily, weekly newspaper, internet news, TV news, etc. What is the difference of media's expression on various different types of

gambling? How has public view on gambling industry changed based on the increase in media's interest? For solving these questions, we discussed on the perception on gambling industry through semantic network analysis using the frequency of news reporting and the degree of public's reaction(public's interest).

II. PREVIOUS STUDIES & THEORETICAL BACKGROUND

Study on the recognition or tendency through semantic network analysis has been increased in the field of journalism. In case study using semantic network analysis on social network service, there was a study done on tourist's using SNS data[6] to see how people view on the topic of reunification of North and South Korea[10][11] and a study on content analysis about Literary therapy using text mining and opinion mining[2]. There was a study on the frequency and text network analysis of the papers in library information science during 7 years since 2004[1]. The study on perception through news analysis was a study on the changing perceptions of tourism in Korea through newspaper analysis [8]. There were also three papers on differences of contents and tones of arguments among newspapers using text mining analysis[1], frame transition of personal information leakage from 1984 to 2014 using social network analysis approach[5], and a semantic network analysis of the newspaper articles on big data[3].

As we can see the review of precedent papers, semantic network analysis has been conducted in various areas. Semantic network analysis using newspaper articles frequently has been used to understand the framing of mass media on special subjects. Semantic network analysis is the method for applying social network analysis to the message on communication[12]. Social network analysis is the method to unearth the relation between actors who has the force to derive social situation. By the same principle, we would like to understand the implications of the documents to be analyzed by understanding the structural relationship between the words that are the components of the message through semantic analysis. We also try to find cognition about gambling industry using these network analysis and semantic analysis. We also try to understand cognition on gambling industry using these network analysis and semantic analysis.

III. RESEARCH DESIGN

Major research questions to be confirmed through big data analysis of news related to gambling are as follows.

■ How is the main content of the news related to gambling industry differentiated by the fields of gambling industry?

■ What are the characteristics of framing(frequency, intention) and network(connectivity, centrality) according to the topic of the news on gambling industry?

■ What are the characteristics of framing and network of news when the contents of news related to gambling industry has been formed as public agenda?

The content that is reported on the mass media, especially news, depends very much on the framing that the media creates. But news is re-reported, deepened and spread according to public reaction to news. In other words, news is initially framed by a small number of major media companies, but public reaction makes news solid and spread. This has a difference in centrality and connectivity in the social network theory. With this premise, we tried to identify public perception through news related to gambling industry.

The subject of analysis is 33,749,890 news items from 1990 to 2015 searched by keyword related to gambling industry from database of Mediagaon that accumulated articles of 61 media companies for 35 years. Once searching and selecting based on major game type in Korean gambling industry, the title of news retrieved by lottery(18,963), sports betting(2,277), horse racing(10,720) and casino(11,189) were analyzed after data cleansing except news unrelated to gambling industry. Looking at the number of articles searched by the type of gambling industry by period, lotto and sports betting had few or no article, and even in the case of horse racing and casino, there were not much number of articles in the 1990s. So the news before 1990s was excluded from the analysis. The analysis procedure is as follows:

1. Morpheme analysis of the collected news using our own java program.
2. Obtaining the co-occurrence matrix of words for each article to visualize the refined words (workbook making) using our own php program.
3. Network analysis to derive between centrality from connectivity using NodeXL.
4. Visualizing and performing semantic network analysis to identify cluster of words using NodeXL.

We excluded keywords which has the frequency of occurrence less than two in whole newspaper articles for visualization of more refined data. News were initially reported by a small number of major media companies. However public's reaction has affected how news became solid and spread. In other words, the frequency with which news is reported on a particular topic is largely determined by media's intention[4]. And the centrality that indicates how much news has been linked to other news indicates the degree to which the news has been re-quoted and spread by public since it was first reported. This means that existing news is spread and solid is related to the degree of public reaction to the news. In this study, we designed the analysis framework by reinterpreting the number of times being reported and the degree of connectivity of the news as the intended framing and naturally reacted public perception. The analysis framework of this study is shown in <Figure 1>.
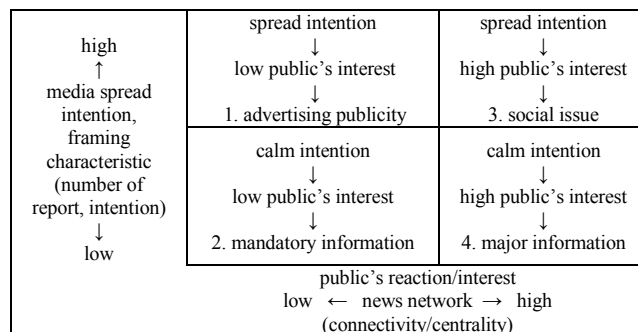

Figure 1. Analysis Famework

The following four categories can be classified by the correlation between degree of news coverage and spread of news. 1. Advertising publicity, 2. mandatory information, 3. social issue, 4. major information. 1. Advertising publicity is likely to be the intended framing information that media company has frequently distributed news for the purpose of spreading the information but public's interest is low. 2. mandatory information that has to be compulsorily reported to a small number of people because media has no will to spread and public is less interested. 3. Social issue is likely to be popular information that news is continually spread because media is willing to spread information and public reaction is high. Lastly, 4. Major information is likely to be an important issue in society. This is because although media company has low intention to report the news, rather, public is more interested in it and even if it is reported less at the beginning, it is spread and cited in the end. In particular, the last 4. major information is likely to be a major reform subjects requiring change in society. Based on this premise, we will identify public perception and find major policy issue about the gambling industry through semantic analysis.

IV. RESULTS OF NEWS ANALYSIS

As described in the research design, the frequency and centrality for each subject indicate the degree of news report and public reaction. In other words, the high frequency of key word means there are many cases reported by the media company. First, we performed the semantic network analysis for each gambling industry. Second, we expressed the group in network graph using intention of the media company. Third, we expressed the link and the connectivity using the relation how news has been spread among other news.

There is a lot of news in the lottery field, and the spread of news is also high afterwards<Figure 2>. The subjects (number of reports) of news that have been much reported are related to the winning result such as win(5,039), number(3,851) and bonus(1,365). The news of highly spread subject(connection centrality) was mostly hopeful contents such as jackpot(28,512), great spot(6,871) and luck(3,637) and it was the field with the greatest number of reports. This implies that the intention of the media company to spread relevant information was high and the public has a great interest in the news. This '**social issue**' news was formed as a '**public agenda**' as follows. Korean government enacted The Lottery and Lottery Fund Act and established "The

Lottery Commission" for the first time in 2004. This was due to public's strong interest since the first issuing of Lotto in December 2002. This resulted in the information of the legal system and the lottery organization for lottery in 2004. This case was the first example that the gambling industry became the major task of public policy in South Korea. This shows that the lottery industry has high popularity and became a major policy area.
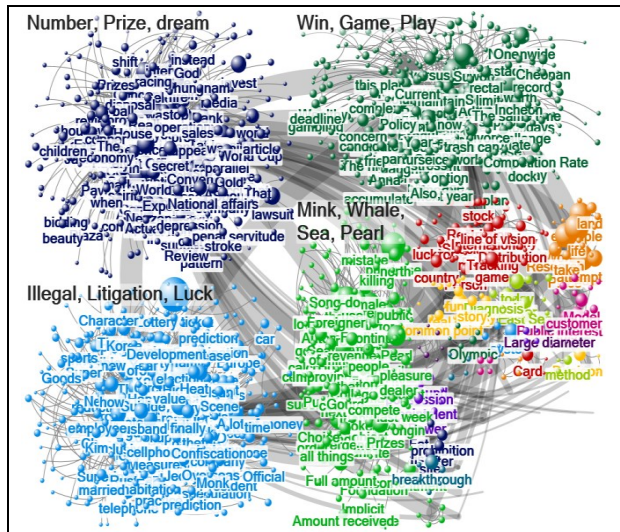

Figure 2.   Lottery related News in the form of Social Issue

In the case of sports betting, the number of news reported is very small compared to other gambling industries<Figure 3>. The subjects of news have been reported are related to the business scandal such as illegal(296), issue(178) and Orion(152). Spread news(connectivity) were news related to manipulation of sports betting results such as football(5,374), detection(1,363), detention(48) and woman(47). It was the field with the smallest number of reports in the gambling industry. The main news was related to Sports Toto's bid business related to a small number of stakeholders. This means that there is an intention to prevent spread of related news in the first place. On the other hand, the degree of spread is relatively small compared to other news as it is small news coverage. This kind of news can be understood in the form of 'mandatory information' framing to minimize the news that must be disclosed to the public. It can also be interpreted that a company that operates sports betting is well controlled by media. The sales of sports betting are higher than the lottery, but the public is less interested. The reason is that lottery tickets are bought by a lot of people, while sports betting is done by a small number of people. In other words, players of lottery game are distributed widely and bet small money for one game. But, players of sports betting are concentrated in narrow group and bet big money for a lot of games by high gambling characteristics.
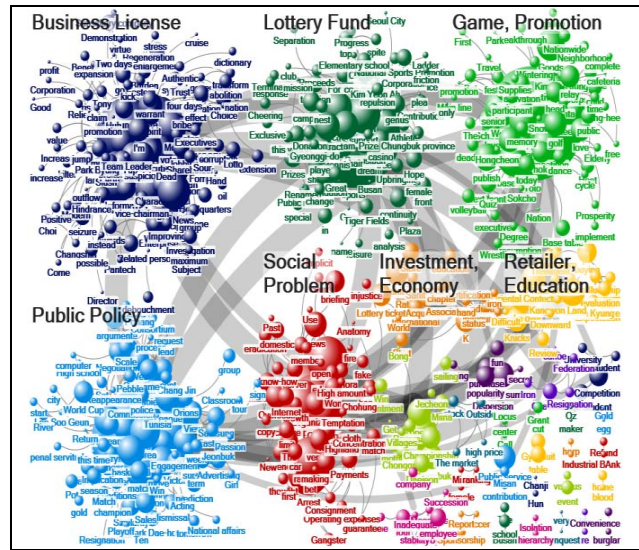

Figure 3.   Sports Betting related News in the form of Mandatory Info.

The number of news reported is high but the degree of spread is relatively low in the horse racing filed<Figure 4>. The major subjects are related to the racetrack facilities such as online-racetrack(2,087), work(1,198), racing(1,035) and Busan(1,009, the provincial city which established new race track). The news of highly spread subject(connection centrality) was related to specialty such as jockey(23,233), horse(9,744) and trainer(4,975). This is the intention of media to provide specific information such as publicity or advertisement to the public. However public reaction was not so high for such publicity information. So the news about horse racing stayed at the publicity level. This phenomenon means that horse racing is a special industry targeting a specific customer.
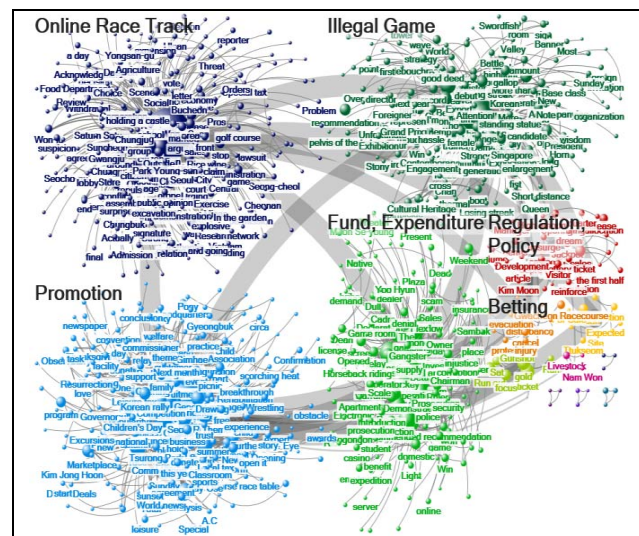

Figure 4.   Horse Racing related News in the form of Advertising Publicity

The number of news reported is relatively low but the spread of specific news is high about casino<Figure 5>. The number of reports is high in the order of gambling place(1,094), Kangwon Land(700, casino company) and foreigners(638). The news about Macau(28,935), local access(41,116) and local permission(15,075) was that connection centrality is high. In other words, media company wants to dismiss the news but became a social issue because public has high interest in the specific issue. This can be understood that a company operating a casino can control the media well but cannot control public's interest.
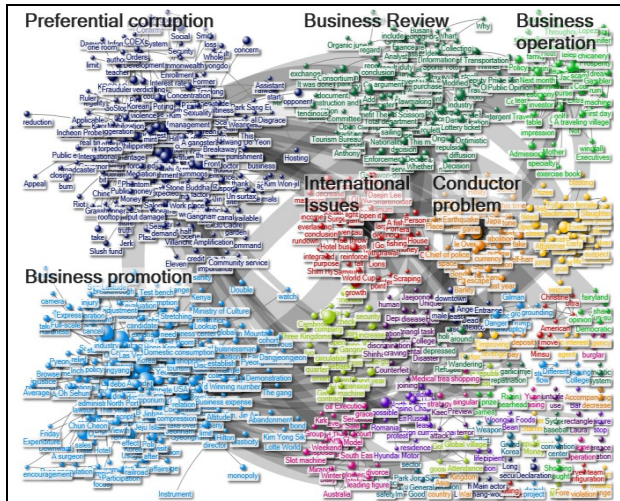

Figure 5.   Casino related News in the form of Major Information

V.   CONCLUSION AND IMPLICATION

<Figure 6> shows the result of news analysis about the gambling by the framing characteristics based on the number of reports and the intention of media and the degree of public's interest based on network connectivity.

spread intention ↑ framing characteristic ↓ dismissal	1. advertising publicity i.e.) horse racing : racing, park	3. social issue → public agenda i.e.) lottery : win, jackpot
	2. mandatory information i.e.) sports betting : manipulation, Orion	4. major information i.e.) casino : gambling, illegality
low ← news spread (public's interest) → high		

Figure 6.   News Framing and Connectivity by Gambling Industry

The analysis of news related to gambling industry is divided into following four categories. 1. advertising publicity, 2. mandatory information, 3. social issue(public agenda), 4. major information. Among them, social issue is the news that is set to public agenda. In other words, news could be formed as a public policy when media's intention is strong and public's reaction is high. It is similar case about consolidation process of agenda setting when Korea Lottery Association was formed in 2006 with strong government's intention and high publics' interest[9]. In addition, we can understand that company about sports bets and casinos have

controlled media and news well because they are an operator under the Ministry of Culture, Sport and Tourism, which is the government agency that supervises media companies. However no one could control public's interest on gambling and lottery industry. And there occurs the difference in public's interest on news by the degree of player's distribution. The public shows high interest in news related to lottery or casino which has wide player's distribution. On the other hand, the public shows lower interest in the news related to sports betting or horse racing because player's distribution is narrow.

In this paper, we analyzed the difference and change in the media reporting framing and public perception of the gambling industry based on news for the last 30 years. In addition, we could explore the policy characteristics and reform subjects by analyzing the frequency and connectivity of articles with framing and public's interest based on the results of semantic network analysis. There was a lot of research on the perception change through semantic network analysis in the existing news analysis. However, this paper is the first attempt to study the setting conditions of public agenda by reinterpreting the frequency and connectivity of network as the importance of information and public reaction. We can also provide to government sector new method to measure the public's cognition about gambling industry instead existing survey.

REFERENCES

[1] Cho, Jane (2011) A Study for Research Area of Library and Information Science by Network Text Analysis, Journal of the Korean Society for Information Management, 28(4): 65-83.

[2] Choi, Kyoung-Ho & Jeong-Hye Park (2015) The Analysis of Public Awareness about Literary Therapy by Utilizing Big Data Analysis, Journal of Digital Convergence, 13(4): 395-404.

[3] Choi, Yoon-jung & Kweon, Sang-Hee (2014) 'A Semantic Network Analysis of the Newspaper Articles on Big data, Journal of Cybercommunication Academic Society, 31(1): 241-286.

[4] Entman, R.M.(1993), Framing: Toward clarification of a fractured paradigm, Journal of Communication, 43: 51-58.

[5] Jeong, Seo Hwa & Hyun Suk Cho (2014) A study on frame transition of personal information leakage, Journal of Digital Convergence, 12(5): 57-68.

[6] Kam, Miah & Min Song. (2012) A Study on Differences of Contents and Tones of Arguments among Newspapers Using Text Mining Analysis, Bibliographic Info: J Intell Inform Syst, 18(3): 53-77.

[7] National Gambling Control Commission (2014) 2014 Gambling Statistics, National Gambling Control Commission, 11-1371045-000037-10.

[8] Min, Kyung Ick & Tae Sook Lee (2012) A Study on the Changing Perceptions of Tourism in Korea through Newspaper Analysis, Journal of Tourism Sciences, 36(10): 81-104.

[9] Moon, HyeJung. (2012a). Study on the Agenda Setting of Lottery Policy. Korean Society and Public Administration, 23(1): 287-316.

[10] Oh, Ick Keun, Tae Sook Lee & Chae Nam Chon (2015) A Study on Awareness of Korea Tourism through Big Data Analysis, Journal of Tourism Sciences, 39(10): 107.

[11] Song, TaeMin (2015) Analysis of the trend of recognition of national unification using social big data, Issue & Focus, 269: 1-8.

[12] Wang, W. & Rada, R.(1998), Structured hypertext with domain semantics, ACM Trans. Inform. Syst., 16: 372-412.

[13] Wasserman, S. & Faust, K.(1994), Social network analysis: methods and applications, Cambridge University Press.