Competition of Mediated Professional Sport in Broadcasting: Applying Latent Dirichlet Allocation with Big Data

Taeyeon Oh, Seoul National University
Namkyeong Jang, Seoul National University
Hyung Keun Koo, University of Pennsylvania
Yukyoum Kim (Advisor), Seoul National University

Marketing Thursday, June 2, 2016 20-minute oral presentation (including questions)
Abstract 2016-040 11:40 AM (Legacy South 3)

Competition in media industry such as TV programs is hard to be explained by traditional economic theories because it does not have a nominal price (or the difference in price is negligible) and it is difficult to observe its characteristics (Goettler & Shachar, 2001). With this perspective, existing literature has focused on the behavior of audiences to analyze the competition in media contents market. There are two distinct categories of the audience; active and passive audience. Repertoire theory provides combined perspectives of active and passive audiences (Shim, 2007). Repertoire in media literature means the set of channels, medium or programs that individuals or households were mainly watching (Heeter, D'Alessio, Greenberg, & MeVoy, 1983). According to Webster (2005), cumulative viewing rating of channels at a prime time is over 90% at the ground-wave broadcasting era. However, it has been decreased in modern media circumstance, which is characterized as multi-channel, -media and -platform. Viewing patterns of aggregated audiences became fragment, but on the individual level, people become more loyal to their favorite channels. That means the viewing patterns converge to some channel repertoires.

Based on repertoire theory, researchers can analyze the competition of channels based on active audience perspective that is explained by use and gratification theory (Shim, 2007). From the U & G theory, audiences seek their needs actively and get gratified by media content according to their needs. The formation of repertoires of the audience can be explained by U & G process (Severin & Tankard, 2000). Specific viewing decision of individuals among their repertoires can be explained by passive audience perspective so-called audience flow (Ehrenberg, Goodhardt, & Barwise, 1990). By analyzing channel repertoires, researchers can better understand how channels (or medium) compete and audiences (or users) make their decision. Previous research studied about repertoires but it focused on mostly channel (Heeter, 1985; Webster, 2005; Webster & Ksiazek, 2012; Ferguson & Perse, 1993; Yuan & Webster, 2006) of medium platforms (Yuan, 2011; Jung, Kim & Chan-Olmsted, 2014; Greer & Ferguson, 2015; Taneja, Webster, & Malthouse, 2012).

To investigate competitions in media contents such as mediated sport, analysis of microscopic repertoires including program repertoire need to be explored but only few research has been conducted. This was because (1) most studies of repertoire have been done by media researchers who were not concerned about microscopic issues such as a certain program and (2) only limited methodologies were applied to construct repertoires, i.e. survey based on memory. In this study, we attempt to provide a new methodology using big data to investigate microscopic repertoires such as program repertoires. This methodology enables the researchers to analyze audience watching the behavior of each program, conduct market segmentation of media audiences and analyze competition structure of sport media in broadcasting market.

Latent Dirichlet allocation (LDA) is a model that allows sets of observations to be explained by underlying unobserved groups with similar traits (Blei, Ng & Jordan, 2003) originated from natural language processing field. In the natural language processing perspective, words from documents are a mixture of various topics that are unknown to researchers and LDA technique can help investigate how observations can be categorized into some topics. Also, it can provide the probability that each document are related to the certain topic. Tirunillai and Tellis (2014) suggests how LDA analysis using big data can be applied in the marketing studies such as market segmentation, brand map and dimensions of reactions. We applied LDA to sport media field directly matching words to programs, documents to audiences and topic to program repertoire segmentations.

This study used the viewing log of ABC Nielsen's TV rate research panels in Korea. This data were recorded that who watched which program, and when. We investigated the data from April 1st at 2013 to June 30th. While raw
data recorded viewing the behavior of audience in seconds, we regarded viewing a program was stayed at the channel over 5 minutes to distinguish watching and searching programs. 1577 panels and 5669 programs were selected to the research data. We constructed viewing matrix with rows as program and columns as the audience. After conducting LDA to viewing log matrix, we allocated audiences to each group that showed the highest probability. Descriptive statistics of demographic information such as age, sex, income, occupation, marriage status and education status for audiences was done after clustering.

From the LDA, we derived 8 repertoire patterns that were operationally defined as (1) multi-program audience group, (2) news-comedy major group, (3) comedy major group, (4) comedy-children group, (5) comedy-sports major group, (6) sports major group, (7) news major group and (8) news-sports major group. We found three segments that were watching sports media, but each segment showed different competition structure. First, events such as baseball, soccer and golf were competing in sports major group. In the comedy-sports major group, a major opponent of sports program was a comedy show, and drama was minor. At last, the news was the main rival to sports shows to audiences of news-sports group. Each group has distinct characteristics. Sports major group mostly consists of middle-aged males as a comedy-sports group was younger and single while the news-sports group was the oldest and married.

This study applied LDA model to segmentation of audiences and program clustering to analyze competition of mediated sports in TV contents. Sports spectators are divided into three groups, and each group has specific characteristics. Our finding indicates that for younger generations, sports are competing with comedy shows while it competes with news and daily drama for aged generations. Peterson and Raney (2008) suggested that sports were similar to drama because of its suspense and Sayre and King (2003) also maintained the similarity of sports and heroic fictions. However, the results of this study show that the main opponent of the sports program is news for elders and comedy for young generations. The viewing patterns are fragmented but polarized similarly to the results of Webster (2006).

This research contributes to the literature of the competition of contents. Previous studies investigated competition of media contents using survey or econometric models to verify substitution or complementation of contents. We suggested applied a methodology to analyze competition in the repertoire perspective. This study also contributes marketing by suggesting competition structure and new segmentation methods in sports media industry. For strategic perspective, this study provides a better understanding of sports media market and consumers. We used Nielsen panel for research because of personal information protection, but broadcasting stations has its audience data and their viewing log of programs. By conducting suggested methodology, more accurate strategic analysis can be available.