ROAD SAFETY POLICY ISSUE EXPLORATION METHODOLOGY ON LIMITED SUBJECTS USING LDA TOPIC MODEL

Chang-Seok Oh
The Board of Audit and Inspection of Korea
112 Bukchonro, Jongno-gu Author address, Seoul, Korea
Phone: + 82-2-2011-3041 E-mail: csoh@bai.go.kr

Co-authors; Yongtaeck Lee, The Board of Audit and Inspection of Korea

ABSTRACT

Study objective: The media represents the right of the people to know, and at the same time collects and forms public opinion about policy demands. For these reasons, media reports cannot be overlooked as a means of monitoring public opinion when it comes to policy formulation. Therefore, this study has developed a technique to search and evaluate policy issues based on quantification of media press releases on the topic of road safety.

Data sources: Data was consisted of media press releases searched in the news library of Naver, Korea’s leading Internet portal. Crawling technique was employed to collect data.

Selection criteria: The search covered a time period of about 23 years, from 1990 to May of 2016. The analysis section was divided into three sections, the ‘3rd period’ referring to the last three years from May 2014 to May 2015, the ‘2nd period’ in which contains the inflection point for quantitative growth of civilian security related media reports from 2002 to 2013, and ‘1st period' referring to the time prior to that.

Data screening and refinement: The initially collected data were subjected to certain screening and refining process. First, we created a limited keyword for the topic to be analyzed and collected press release data through crawling. In addition, we selected data that include both keywords and negative opinions based on one sentence of the press release.

Analysis method: Media data is comprised of unstandardized information in text format. Therefore, in this study, topic analysis technique as a form of text mining was used to analyze unstandardized data. Topic analysis is a method of clustering documents with similar themes by applying a text mining model to collected documents crawled based on search keywords. Through subject analysis, ‘topic keywords’ and ‘topic contribution’ are calculated, which can be used as quantitative indicators to help judge policy issues. The LDA model was used for the specific topic subject.

Results: In this study, we introduced the 'TF-IDF' index in order to examine the changes of keywords in the media reports. This index indicates how much and evenly the keyword was used in the collected press releases. The larger the value of keyword, the more frequently it has been mentioned in various media. Based on the TF-IDF index, the weighted words were classified as 'high speed', 'construction', 'safety', 'traffic' in the 1st period(1990-2002), 'high Speed', 'safety', 'transportation', 'facilities' 'construction’ in the 2nd period(2003-2013), and lastly, 'speed limit,' ‘speed bump,’ ‘address,’ ‘road signs' etc. in the 3rd period(2014—May 2016). Respectively, the safety of highway facilities has constantly been raised as issues, and concerns on safety problems related to speed bumps and traffic signs are getting elevated more recently in the third period.

The policy issues related to the road traffic safety results, analyzed by the period via topic analysis on the press release, are as follows. The most important issues related to road traffic safety were related to
the 'inconvenience to charging high pass fare (2001)' and 'inter-Korean railway and road diplomacy hardships (2002)' during the 1st period (1993-2002). In the 2nd period (2003-2013), concerns on the lack and unmeasured sizes of speed bumps, such as "no speeding brakes in front of elementary school, no braking brakes (2005)" and "speed bumps ignoring the standard measurements (2012)" have been raised. In the 3rd period (2014-May 2016), issues related to "dissatisfaction with road infrastructure in Seoul (2015)" and "risk of speeding bumps (2015)" were raised, which makes it noteworthy that the safety issue of the speed bump is kept mentioned as it did in the second period.

Keywords: Road safety policy issue analysis, text mining, topic model, LDA model, Issue Exploration