

SUMMARY

Keywords: Public Opinion Monitoring, Spatiotemporal Analysis, Unstructured Big Data

Recently, a lot of efforts have been made to identify policy demands hidden in public opinions created through the Internet social media (ISC). Nonetheless, methodological challenges still remain to differentiate delicate changes in the trends of public opinions across space and time. This study aims to overcome such challenges by developing a new model for spatiotemporal analysis of unstructured big data from ISC and directing their uses in public policies.

Conceptually, our model seeks to combine the applications of text mining, social network analysis, spatiotemporal analysis to unstructured big data from ISC so that it can become a data-centric method for public opinion monitoring. The ultimate goal of this model is to support timely identification of policy demands tailored to the needs of the public and regions by enabling the continuous analysis and monitoring of spatiotemporal patterns in public opinions from ISC.

Methodologically, our model consists of three stages for the collection, preprocessing, and spatiotemporal analysis of unstructured big data. The first stage of our model involves the incessant gathering, storage, and management of messages that include discussions of the public on ISC. Oftentimes, this stage

necessitates the use of open APIs, and free or commercial software agents for the crawling of documents from ISC. The second stage of our model involves three tasks: 1) removal of advertisements, swearwords, and redundant messages; 2) linking messages to their associated locations/places (geocoding or spatialization); and 3) conversion of individual messages into a set of substantive terms. The final stage of our model focuses on exploring spatiotemporal patterns in the contents and creators of public opinions through the mixed applications of spatiotemporal analysis, time series data analysis, text mining, social network analysis, and etc. In specific, the contents of public opinions can be understood by analyzing spatiotemporal variations in the locations of message posting, spatio-temporal clusters of term usage, and topical and emotional trends hidden in a corpora of messages/documents. Meanwhile, the analyses of the locations and social networks of message authors could reveal the spatiotemporal distributions and interactions among the creators of public opinions.

The model proposed in this study contributes to the literature of public opinion research by providing a new analytical framework that can detect and visualize spatiotemporal patterns in the contents of public opinions as well as interactions among the creators of such opinions. For a policy standpoint, the proposed model is expected to develop into a software infrastructure for the shared use by various public institutions for the purpose of public opinion monitoring.