Development of a Prediction Engine for Scientific Technology Trends

Ju Seop Park, Soon-Goo Hong
Department of Management Information Systems
Dong-A University
Bumin Campus, 225 Gudeok-ro, Seo-gu, Busan
Korea
juseop60@naver.com, shong@dau.ac.kr

Abstract: The Delphi method is typically used to predict promising technologies that may be invented in the future, but this approach is a difficult process that entails considerable time and cost. An alternative is latent Dirichlet allocation (LDA), which is a topic modeling method. In this research, LDA was used to develop a fast and low-cost prediction engine for scientific technology trends. The proposed engine features the use of text mining that is targeted toward the abstracts of patent documents. Aside from LDA, required fundamentals were used as bases for the design of the proposed engine. The results are expected to contribute to future research on scientific technology trends. The developed engine can also be used as an economical, automated prediction system in other related fields.

Key-Words: development of prediction systems, scientific technology trends, technological prediction, text mining, topic modeling, analysis of technological trends

1 Introduction

As the competition to develop cutting-edge technologies accelerates, corporations are expected to establish and promote an increasing number of innovations to secure a leading position in future markets (Jo et al., 2009). Accordingly, practitioners and researchers have begun paying attention to trends or expectations about future technologies. In the Korean context, scientific technology prediction was initiated in 1994 and has been conducted every 5 years since. To this end, Korean experts in scientific technologies typically use the Delphi method, but this approach involves high cost and considerable time. Given that Korea treats the provision of fundamental materials for scientific explorations as a meaningful goal, the government established a scientific technology policy. Such policy, however, is difficult to apply in private industries. Predicting the circumstances that may confront corporations in the future and selecting promising technologies necessitate a focused and well-chosen long-term strategy. This requirement prompted the adoption of latent Dirichlet allocation (LDA), which is a relatively inexpensive approach to predicting the trends associated with scientific technologies.

On the basis of the above-mentioned considerations, this research developed a prediction engine for scientific technology trends using the LDA topic modeling method. To verify the proposed engine, the abstracts of US artificial intelligence (AI) patent documents were used in the prediction of scientific technology trends. To confirm the prediction results, the frequencies with which keywords arise with respect to relevant topics were determined.

2 Theoretical Background

2.1 Research on Technological Trend Prediction

Technology involves the application of scientific theory to the real world and enables the processing of various components to ensure usefulness to human life. Technology is generally employed in industries or scientific fields, where the development of scientific technology is a fundamental step in the establishment of industrial technology.

Technological prediction means the forecasting of the future status of a specific technological area (Jeon et al., 2014). Such task can provide information that is important to the formulation of a national investment strategy for research and development (R&D). Technological prediction also enables active response to changes in the circumstances that surround scientific technologies and agreement among experts who are in charge of establishing scientific technology policies.

Furthermore, technological prediction contributes to the development of a reasonable basis for organizing the R&D initiatives of governments (Ahn et al., 2003).

Methods of technological prediction can be classified into qualitative approaches based on intuition and quantitative methods based on data. A qualitative method assumes that various futures can happen and that experts or practitioners are qualitatively expected to respond to such situations. In this approach, therefore, focus is directed toward improving crisis management. A quantitative method involves prediction based on quantitative analysis, with emphasis placed on supporting a reasonable decision-making process in the current period (Jeong et al., 2006). Examples of qualitative prediction methods are the Delphi method, brainstorming, the use of expert panels, and the scenario method. Examples of quantitative prediction approaches include the trend of patent analysis, the use of growth curves and system dynamics, and cross-impact analysis.

2.2 Research on Technological Trend Prediction using Topic Modeling

LDA, which is a most widely used topic modeling method, was proposed by Blei et al. (2003). It is a static algorithm used to discover subjects that are potentially hidden within extensive and atypical groups of documents. LDA draws attention to the benefits of text mining, facilitates the identification of potential topics from various documents, and enables effective analysis through suggestions regarding the weight of topics about entire documents. Previous studies based on this technique can largely be divided into trend analyses (Kim, 2016 ; Kim et al., 2016), technical studies, such as the extraction of technical topics (Kim et al., 2017 ; Jeong et al., 2015), and social studies, such as the derivation of social trends (Roh, 2016 ; Lee et al., 2014).

3 Design of the Proposed Prediction Engine

The framework that underlies the proposed prediction engine, which operates on the basis of text mining, comprises a DB system, an LDA system, and a technological trend analysis system (Fig. 1). A DB is extracted from patent documents, journals, and news articles that consist of data useful for the prediction of scientific technology trends. The LDA system is the S/W that can provide data for examination in the technological trend analysis stage. Topic modeling can be divided roughly into preprocessing and topic classification.

The technological trend analysis system analyzes the weight of each period and technology for a thorough understanding of core technology. “Hot” and “cold” technologies are extracted on the basis of the weight analysis for each year in an entire period. The results are then visualized.

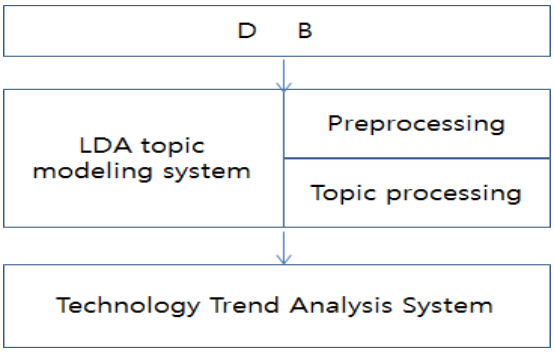

Fig. 1. Framework of the Prediction System

4 Application of the Proposed Engine

For scientific technology prediction using the proposed system, the abstracts of US AI patent documents, which are widely applied in the analysis or prediction of scientific technology trends, were used. The adoption of the key phrase “artificial intelligence” yielded 13,618 patent data documents, including information on patent registration date, patent application data, patent name, and patent abstract, by year in the US patent DB between 2002 and 2016. Targeting the abstracts of the collected patent documents, the preprocessing step, which involves tokenization, stop word removal, and stemming, was implemented. Then, the topics associated with AI technology and corresponding theta values were extracted via LDA. The theta values were used to calculate the proportion of topics arising in a given year and the rake value weights of the topics. Finally, the technological flow of the topics was determined through regression analysis.

According to the trend analysis for the AI topic, six hot technologies and six cold technologies were identified. To validate the 12 cases of prediction, the frequencies with which related keywords arose during the past three years (2014-2016) were determined.

5 Conclusion

In this study, a scientific technology trend forecasting engine was developed. To this end, the study also made an academic contribution to facilitating follow-up studies using text-mining techniques. Moreover, the developed forecasting engine will likely be used in actual field operations. The results of a trend analysis on the topic of artificial intelligence derived six hot technologies and six cold technologies. The six hot technologies were computer systems, multimedia systems, company architectures, wireless technologies, games, and servers. The six cold technologies were power systems, LED light systems, sound processing, JAVA development systems, virtual reality, and picture preprocessing. The recent three-year (2014–2016) trends in the frequency of words associated with the 12 technologies under the topic of AI were analyzed to test the validity of the 12 forecasted technologies. The forecast results from the recent 15-year trend analysis and the recent three-year frequency analysis of associated words within the topic were compared to measure how well they matched. As a result, eight out of 11 technologies coincided (one technology could not be tested).

References:

[1] Ahn, D. H., Shin, T. Y., Mun, M. J. and Kim, H. S., Future Socio-Economic Issues and Needs for Technology Foresight, Science and Technology Policy Insititute, Report, 2003.

[2] Blei, D. M., Ng, A. Y. and Jordan, M. I., "Latent dirichlet allocation", The Journal of machine Learning research, Vol. 3, pp. 993-1022, 2003

[3] Jo, B. S., Ji, K. Y., Kim, Y. J. and Lee, B. G., Future Technology Forecast Process, ETRI Planning Report, 2009.

[4] Jeong, S. Y., Nam, S. I., Hong, S. and Han, C. H., "Future Technology Foresight for an Enterprise : Methodology and Case", The Journal of Society for e-Business Studies, Vol. 11, No. 1, pp. 69-82, 2006.

[5] Jeon, S. H., Park, S. S. and Jang, D. S., Patent Analysis and Technology Forecasting, Kyowoosa, 2014.

[6] Kim, J. H., "Research Trends Analysis for 'Internet of Things'Based on Topic Modeling and Network Analysis", Seoul National University of Science and Technology Master's thesis, 2016.

[7] Kim, S. K., and Jang, S. Y., "A Study on the Research Trends in Domestic Industrial and Management Engineer", Journal of the Korea management engineers society, Vol. 21, No. 3, pp. 71-95, 2016.

[8] Kim, M. A., and Suh, C. K., "SCM Patent Analysis Using Topic Modeling: 1997~2016", Journal of the Korean Society of Supply Chain Management, Vol. 17, No. 2, pp. 19-29, 2017.

[9] Suh, S. H., and Lee, H. Y., "Fintech trend analysis using topic modeling of BM patents", Proceedings of the Fall Conference of the Korean Institute of Industrial Engineers, Vol. 2015, No. 11, pp. 471-480, 2015.

[10] Noh, B. J., Suh, J. S., Lee, J. U., Park, D. H., and Chung, Y. H., "Keyword Network Based Repercussion Effect Analysis of Foot-and-Mouth Disease Using Online News", Journal of Korean Institute of Information Technology, Vol. 14, No. 9, pp. 143-152, 2016.

[11] Lee, S. Y., and Lee, K. M., "Trend Extraction using Topic Model Based on Reply Graph", Proceedings of the Conference of the Korean Institute of intelligent systems, Vol. 24, No. 2, pp. 99-100, 2014.