

A New Forecasting System using the Latent Dirichlet Allocation (LDA) Topic Modeling Technique

JU SEOP PARK, NA RANG KIM, HYUNG-RIM CHOI, EUNJUNG HAN

Department of Management Information Systems

Dong-A University

Bumin Campus, 225 Gudeok-ro, Seo-gu, Busan

SOUTH KOREA

juseop60@naver.com, whitecoral@hanmail.net, hrchoi@dau.ac.kr, ejhan.biz@gmail.com

Abstract: Although the Delphi technique is often used to forecast promising future technologies, this method is difficult, time-consuming, and costly. As an alternative, the Latent Dirichlet Allocation (LDA) topic modeling technique can be used. Therefore, this study aimed to develop a science and technology trend forecasting system using the LDA topic modeling technique as a form of text mining. An empirical analysis of 13,618 abstracts regarding U.S. artificial intelligence (AI)-related patents was conducted, and the results of the analysis were verified based on changes in the frequency of related words within the AI topics. The trend analysis of the AI topics resulted in six hot technologies and six cold technologies. The results of the verification showed that 8 out of the 11 technologies matched (1 technology could not be verified). This study provides a foundation for engine design by helping develop engines that enable simple and low-cost technology forecasting and by suggesting an appropriate topic modeling technique. The study also makes an academic contribution by encouraging follow-up studies. Moreover, the developed forecasting system may be used as an automated forecasting engine to conduct tasks related to regional innovation.

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

1 Introduction

Amid today's increasing competition for the development of innovative technologies, corporate demands for predicting and developing promising future technologies are expected to grow [1]. The increase in such demands has heightened interest in the forecasting of future technology trends. In South Korea, science and technology foresight has been conducted every five years since 1994. Although the Delphi technique, which involves technology experts, is often employed to conduct technology forecasting, this method can be expensive and time-consuming. Moreover, although the Delphi technique can be valuable for national projects that aim to provide basic data to establish science and technology policies, it is difficult to apply the technique in the private sector. Today, it is necessary to establish long-term strategies for policies and businesses through selection and concentration by forecasting future corporate environments and selecting promising technologies. To this end, text mining emerges as an easy-to-use and low-cost technology forecasting method. Text mining extracts meaningful information using natural language processing technology and various analysis techniques. This study aims to overcome

the weaknesses of the Delphi technique by using topic modeling as a text mining technique. Topic models are a type of statistical model for automatically discovering topics that occur in a collection of documents using algorithms. Topic modeling can overcome the limitations of conventional qualitative analyses and identify the hidden topics in many documents. The most typical topic modeling technique is Latent Dirichlet Allocation (LDA), which is actively used in technology forecasting research in areas such as technology management, library and information science, and computer science [2]. While there are several methods for forecasting science and technology trends, using these methods involves several difficulties. Therefore, the development of a forecasting engine that is easily accessible and convenient for users will facilitate future technology forecasting research.

The purpose of this study is to develop a science and technology trend forecasting engine using the LDA topic modeling technique. To this end, an empirical analysis of abstracts regarding U.S. artificial intelligence (AI) patents, which are widely used to forecast and analyze science and technology trends, was conducted, and the results of the

analysis were verified based on changes in the frequency of related words within the AI topics

2 Theoretical Background

2.1 Technology trend forecasting research

Science and industry technologies are a means of applying scientific theories to actual fields and processing them to be useful for human life. The development of science and technology is a foundation for the development of every industrial technology. Technology forecasting refers to forecasting the future conditions of a specific technical field [3].

Technology forecasting can provide important information to establish a country's research and development (R&D) investment strategies. Technology forecasting enables the country to actively respond to environmental changes in science and technology, experts to reach an agreement in the process of establishing science and technology policies, and the government to establish a rational basis for R&D budget planning [4]. In addition, technology forecasting can help the country to proactively acquire core technologies by focusing on investment in promising future technologies. Today, scientific and technological competition is becoming increasingly important in terms of securing the capacity for innovation to continue social development and achieving the growth of the national economy. Moreover, while limited resources in various areas may need to be effectively used through science and technology, the outcomes of scientific and technological initiatives are uncertain. Therefore, measures to select promising technologies should be prepared at the strategic level [5].

Technology forecasting methods can be divided into intuition-based qualitative methods and data-based quantitative methods. The qualitative methods focus on responding to future circumstances by assuming the occurrence of various circumstances in the future, projecting the future qualitatively, and improving the ability for risk management, whereas the quantitative methods focus on providing the conditions that enable reasonable decision making in the present circumstances through future forecasting based on measurement techniques [6]. The qualitative methods include Delphi, brainstorming, and scenario methods, and the quantitative methods include patent trend analysis, system dynamics, and cross-impact analysis.

2.2 Technology forecasting research using topic modeling

Suggested by Blei et al. [7] as a statistical algorithm to discover potential topics in an extensive and atypical collection of documents, LDA is the simplest topic modeling technique and is universally used. LDA topic modeling, which is drawing attention in the field of text mining, allows for the discovery of the hidden topics within several documents and performs an effective forecasting analysis by suggesting the proportions of the hidden topics in the entire set of documents. Previous studies that have been conducted using LDA include trend analyses ([8], [9]); technical studies, such as the extraction of technical topics ([10], [11]); and social studies, such as the derivation of social trends ([12], [13]).

3 Design of the science and technology trend forecasting engine

3.1 Definition of the framework

As shown in Fig.1, the framework for the present study's science and technology trend forecasting system using a text mining technique is a minimum configuration system that includes documents, preprocessing, and topic classification, which are necessary to develop a forecasting system to predict science and technology trends. In other words, the establishment of a forecasting system based on an accurate framework will result in the effective forecasting of science and technology trends.

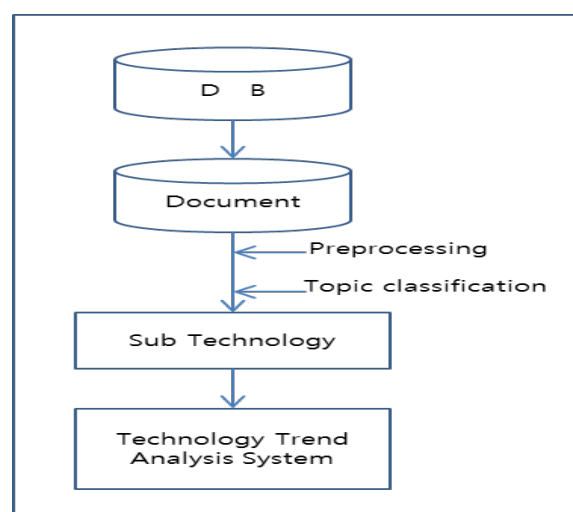


Fig.1 Definition of the framework

3.2 Components

As shown in Fig.2, the framework for a science and technology trend forecasting system using a text mining technique consists of a database (DB), an LDA topic modeling system, and a technology trend analysis system. The DB is extracted from documents that contain data that enable forecasting, such as patent documents, academic journals, and new articles. The LDA topic modeling system provides data that can be analyzed by the technology trend analysis system, which consists of two tasks: preprocessing and topic classification. Preprocessing is a preliminary task that enables effective topic classification and includes the tokenizing, stop-word elimination, and headword modules. The topic classification task includes selecting options (number of topics, number of sampling repetitions, etc.) for the topical modeling package, specifying label names for topics, and receiving and maintaining θ values. The technology trend analysis system analyzes the periodic proportion of each sub-technology to identify core technologies. The system derives and visualizes promising and declining technologies by analyzing annual trends in the proportions of technologies.

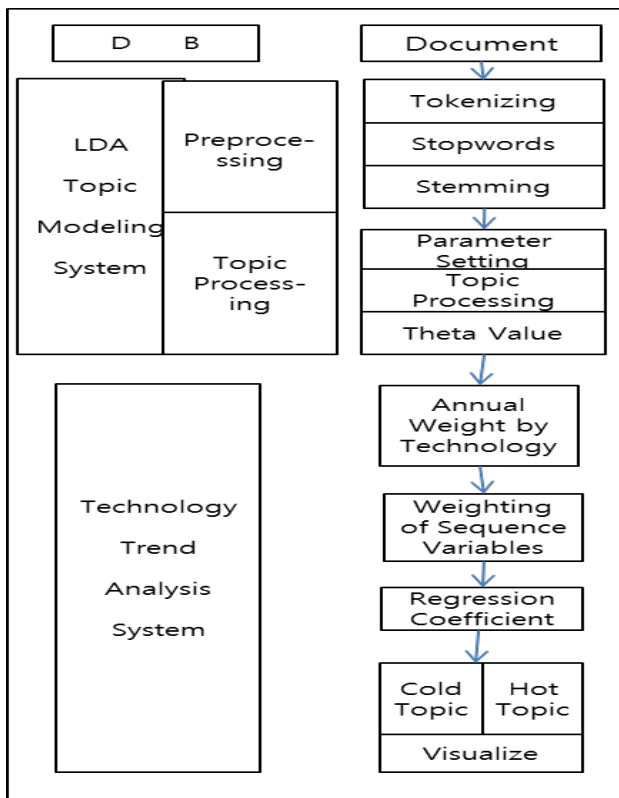


Fig.2 The LDA topic modelling technique for the science and technology trend forecasting system framework

3.2.1 The LDA topic modeling analysis system

3.2.1.1 Preprocessing

The tokenizing module converts a text that is comprised of many sentences into individual words based on the spaces of the text. In doing so, each text is indicated as a collection of multiple words, which facilitates smooth data mining. The articles, prepositions, and special characters that are typically used in English documents can impede data analysis; thus, the stop-word elimination module eliminates them in advance. Stemming refers to the task of removing the endings of words that are used as various parts of speech in English documents and extracting the stems of these words. For example, the stem of the character strings “argued,” “argue,” and “arguing” is “argu.” The framework for developing stemming is known as “Snowball.”

3.2.1.2 Topic preprocessing

After the preprocessing process is completed, the topic analysis is conducted. Basic parameters, including the number of topics, the number of sampling repetitions, and the α and β values, are set. Appropriate numbers of topics and sampling repetitions can be determined based on the levels at which the researcher can effectively perform the analysis of outputs [14].

In this study, the number of topics and the number of sampling repetitions for LDA topic modeling were set at 20 topics and 10,000 times, respectively, using the topic models of R, a statistical analysis program. The α and β values were set at a default value of 0.1. The parameters for LDA topic modeling are presented in Table 1.

Table 1 Content of parameters

| Parameter | Content |
|-----------|---|
| α | This parameter affects the distribution of topics within documents. If the value increases, then the number of topics within the documents decreases. If the value decreases, then the number of topics within the documents increases. If the researcher does not specify it, a basic value is set by the collection of documents. |
| β | This parameter affects the distribution of words within topics. If the value increases, then each word is distributed over a number of topics. If the value decreases, then each word is narrowly distributed over specific topics. |

| | |
|-----------|---|
| Iteration | This parameter indicates the number of sampling repetitions, and sampling is repeated for proper topic modeling. The researcher can decide this value by testing the results while gradually increasing the number of sampling repetitions and selecting the level that enables effective analysis. |
|-----------|---|

3.2.2 The technology trend analysis system

As an output, θ values are produced using the topic models of the R program. The θ values are received in the following format via an Excel file: M (total number of patents)*20 matrix. Patents are listed in the rows, and the respective proportions of the 20 topics are indicated as numerical values in the columns. Based on the date of the patent application, the patents are aligned in ascending order, with the oldest placed in the first row.

3.2.2.1 The proportion of each sub-technology over the entire period

The sum of each proportion in the columns becomes the total proportion of the 20 topics (sub-technologies). The ranks of the 20 topics can be obtained by calculating the percentage of each topic to the sum. In addition, periodic changes in the sub-technologies over the review period can be identified. Specifically, changes in the rank of each sub-technology can be identified by dividing the entire 15-year review period into sub-periods and then calculating the respective proportions of the 20 topics for each period and determining their periodic ranks.

3.2.2.2 Annual proportions of sub-technologies

The entire 15-year review period is broken into 15 one-year sub-periods, and the θ values are classified on an annual basis. The annual proportions of the 20 sub-technologies are calculated. This produces the annual ranks of the 20 sub-technologies in terms of their respective annual proportions.

3.2.2.3 Application of weights to ranking variables

On an annual basis, a weight is applied to the rank of each sub-technology's proportion. For example, the first, second, and twentieth ranks are given weights of 20, 19, and 1, respectively. Because the proportions of the sub-technologies are indicated as relative values (not absolute values), the proportions are first indicated as ranking parameters, and then weights are applied to the parameters.

3.2.2.4 Regression coefficients

In this study, regression coefficients for regression analysis were employed to determine the trend of each technology. A regression analysis was performed using the year as the independent variable and the weight of the annual proportion of each sub-technology as the dependent variable.

3.2.2.5 Promising and declining technologies

In the regression analysis, significant technologies were determined at a statistical significance level of 0.05 or below. The significant technologies were defined as either hot technologies if their regression coefficients were positive (+) or cold technologies if their regression coefficients were negative (-).

3.3 Unified modeling language diagrams

3.3.1 Class diagrams

The class diagram shown in Fig.3 illustrates the structure of the science and technology trend forecasting system using a text mining technique. Once a document is entered, the tokenizing module extracts words (tokens) by dividing the words based on their spaces, commas, and periods. The output of the tokenizing module is documents. Once a document is entered into the stop-word elimination module, the module begins the task of refining words. First, words contained in the English stop-word list are removed. The stop-word list includes 571 words that are frequently used in documents but are unnecessary for search indexes, such as "a," "the," "of," and "an." Next, punctuation marks, numbers, spaces, and special symbols are removed, and the document in which the upper cases of words have been converted to lower cases is transferred to the Snowball module, which performs stemming. The endings of all words within the document are removed, and the stems of these words are extracted; then, the resulting document is transferred to the topic modeling module.

In LDA topic modeling, once the number of topics, the number of sampling repetitions, and the α and β values are set for a topic analysis of the document that underwent the preprocessing step, the topic analysis is performed using the R package topic models. As an interface, the Topic_classify module prints out three files referring to θ values, topic keywords, and π values as output data. The θ file is used as input data for the technology trend analysis.

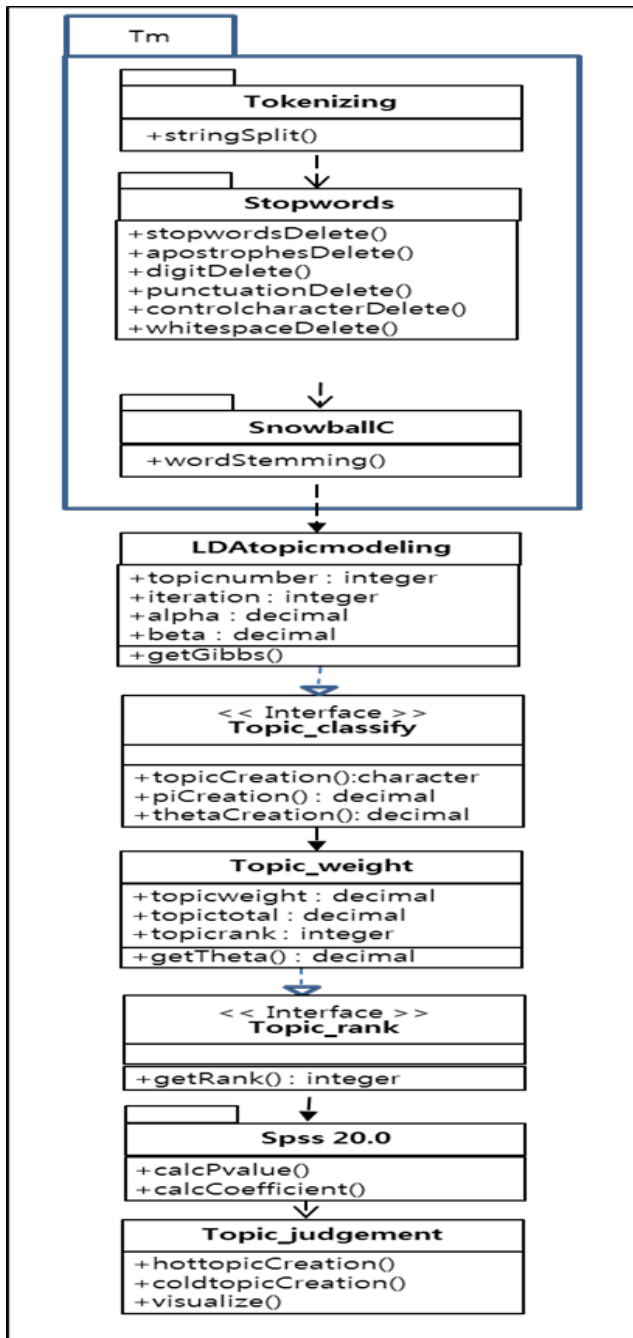


Fig.3. The class diagram of the forecast system

Input data in the Topic_weight module of the technology trend analysis system are received as θ values listed in the M (total number of patents)* N (number of topics) matrix of an Excel file. For example, 20 topics in the rows and 10,000 documents in the columns creates a 10,000*20 matrix. The annual proportion (share) of each topic is obtained by dividing these documents into annual groups and adding up the values for each of the 20 topics on an annual basis. The conversion of each topic's annual proportion into a relative percentage of the total produces the rank of each topic. A high rank is provided with a higher score. For example,

the first and second ranks are given 20 points and 19 points, respectively. Similarly, the twentieth rank is given 1 point. Outputs from the interface Topic_rank module are arranged in a 15(years)*20(number of topics) matrix, and the scores of each topic range from 1 point to 20 points for each year. This matrix is entered into the statistical software SPSS 20.0 and undergoes a regression analysis; then, the β values and significant probabilities of each topic are produced. The Topic_judgment module receives the β values and significant probabilities, determines the hot and cold topics based on the data, and draws diagrams of the selected topics.

3.3.2 Sequence diagrams

A sequence diagram was drawn to represent the order in which the science and technology trend forecasting system using a text mining technique operated. The flow of data processing in the forecasting system is shown in Figure 4. The LDA topic modeling system performs data reprocessing and topic tasks, and the θ file resulting from the tasks is passed on to the technology trend analysis system. The technology trend analysis system classifies the selected technologies into hot and cold technologies by processing the received θ file and schematizes the selected topics.

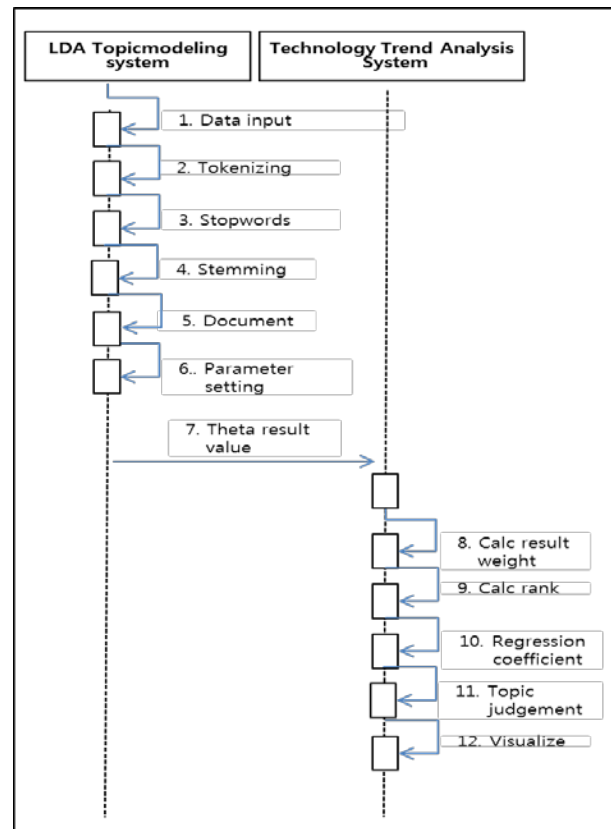


Fig.4 Diagram of the forecasting system sequence

4 Science and technology trend forecasting system

4.1 Summary of the technology trend analysis

In this study, abstracts regarding U.S. AI patents were analyzed for technology forecasting using a science and technology trend forecasting system. Using the keyword “artificial intelligence,” 13,618 patent-related documents published between 2002 and 2016 were extracted from U.S. patent databases. These documents contained data such as the dates of the patent application and registration and the names of patents, as well as abstracts regarding patents. Among the selected patent documents, the abstracts were used to perform preprocessing tasks, such as tokenization, the elimination of stop words, and stemming. The topics and θ values of the AI technologies were then extracted using the LDA topic modeling technique. Using the θ values, the annual proportions of individual topics were calculated, and a weight was applied to the ranking variable of each topic. Technology trends for the topics were then predicted via regression analysis.

4.2 Data collection and preprocessing

A total of 13,618 abstracts regarding U.S. patents published between 2012 and 2016 were collected for science and technology trend forecasting and were classified on an annual basis. Fig.5 presents changes in the annual number of abstracts regarding AI. After the data from the U.S. patents were retrieved from the Korea Intellectual Property Rights Information Service, a patent information retrieval service, patent application dates and numbers, patent registration dates and numbers, patent names, and abstracts were obtained through an online download service. The preprocessing task includes tokenization, which segments multiple texts into individual words based on the spaces of each text; the elimination of stop words used in English documents, such as articles and prepositions; and stemming, which removes the endings of and extracts the stems from words that are used as various parts of speech in English documents.

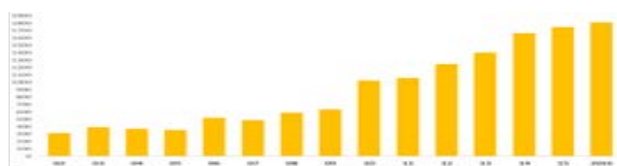


Fig.5 Annual numbers of abstracts regarding U.S. AI patents

4.3 Analysis of AI technology trends

4.3.1 Extraction of AI sub-technologies

To derive the AI sub-technologies, topics were extracted from the abstracts regarding AI patents, which had completed the preprocessing step via an LDA topic modeling program. In this study, 20 topics, 7 related words, 1,000 sampling repetitions, and the parameters α and β were set as default values.

Among the 20 topics extracted as AI sub-technologies, those in the field of basic AI are shown in Table 2, and those in the field of applied AI are shown in Table 3, respectively. Three AI experts selected the study’s topics and removed the words that had markedly low levels of relevance to the topics from the initial selection of related words.

As topics in the field of basic AI, eight technologies were selected: [T1] computer systems, [T3] company architecture, [T6] file operation, [T12] databases, [T14] JAVA development systems, [T17] networks, [18] time, and [T19] servers. As topics in the field of applied AI, 12 technologies were selected: [T2] multimedia, [T4] power systems, [T5] LED light systems, [T7] sound processing, [T8] wireless technology, [T9] healthcare, [T10] object recognition, [T11] robots, [T13] text processing, [T15] games, [T16] virtual reality, and [T20] picture preprocessing.

Table 2 Words related to the eight sub-technologies in the field of basic AI

| [T1] Computer System | [T3] Company Architecture | [T6] File Operation | [T12] Data Base |
|---------------------------------------|---|--|--|
| computer program code method | portion set subset processor | event rule file key set | subject test data feedback |
| [T14] Java Development System | [T17] Network | [T18] Time | [T19] Server |
| data input stream asset | network node traffic address packet | time real item tool risk work | server client web request call internet |

Table 3 Words related to the 12 sub-technologies in the field of applied AI

| | | | |
|---|---|---|--|
| [T2] Multimedia | [T4] Power System | [T5] LED Light System | [T7] Sound Processing |
| content audio context display multimedia | power circuit fluid control line grid | sensor light pattern field phase detector | Signal Input Unit Output Speech Sound Audio |
| [T8] Wireless | [T9] Health Care | [T10] Object Recognition | [T11] Robot |
| wireless access radio station base | patient health treatment person blood | object target region point interest center | Control State Robot Action environment behavior |
| [T13] Text Processing | [T15] Game | [T16] Virtual Reality | [T20] Picture Preprocessing |
| Search Text document Word Answer Section Domain | game virtual display character player screen | video motion camera frame Map Pixel color | Mode Vector Matrix algorithm Graph |

4.3.2 Hot and cold AI technologies

After classifying the abstracts regarding U.S. AI patents on an annual basis, the proportions of the individual technologies were also analyzed. Hot technologies whose values showed upward trends and cold technologies whose values exhibited downward trends were derived from the analysis of trends in the annual proportion of each technology over the entire review period. As the standard for assessing the trends of individual technologies, regression coefficients for regression analysis were employed. A simple regression analysis was performed using the year as the independent variable and the annual proportions of the individual technologies as dependent variables. The statistical significance level was set at 0.05. Each significant technology was determined as either a hot technology if the regression coefficient was positive or a cold technology if the regression coefficient was negative. Based on the analysis results, six hot technologies and six cold technologies were derived (Table 4). The six hot technologies were [T1] computer systems, [T2] multimedia, [T3] company architecture, [T8] wireless technology, [T15] games,

and [T19] servers (Fig.6). The six cold technologies were [T4] power systems, [T5] LED light systems, [T7] sound processing, [T14] JAVA development systems, [T16] virtual reality, and [T20] picture preprocessing (Fig.7). The ambiguous technologies, which could not be clearly classified as hot or cold, were [T6] file operation, [T9] healthcare, [T10] object recognition, [T11] robots, [T12] databases, [T13] text processing, [T17] networks, and [18] time (Fig.8).

The areas that have widely been researched in recent years include autonomous vehicles, medical care, and finance. Autonomous vehicles refer to technologies such as [T1] computer systems, [T2] multimedia, [T8] wireless technology, and [T19] servers. Medical care requires technologies to increase the accuracy of medical diagnoses, such as [T1] computer systems, [T2] multimedia, and [T19] servers. The financial industry is paying close attention to investment and trading systems, personal banking support systems, credit assessment and evaluation systems for loans, and illegal practice detection systems for financial transactions. These four fields require technologies such as [T1] computer systems, [T2] multimedia, and [T19] servers.

Table 4. Regression analysis of the 20 AI technologies

| No | AI topic | Beta | P-value | Hot/Cold |
|----|----------|--------|---------|-----------|
| 1 | [T1] | 0.543 | 0.036 | Hot |
| 2 | [T2] | 0.881 | 0.0 | Hot |
| 3 | [T3] | 0.641 | 0.01 | Hot |
| 4 | [T4] | -0.733 | 0.002 | Cold |
| 5 | [T5] | -0.824 | 0.0 | Cold |
| 6 | [T6] | -0.030 | 0.916 | Ambiguous |
| 7 | [T7] | -0.633 | 0.011 | Cold |
| 8 | [T8] | 0.865 | 0.0 | Hot |
| 9 | [T9] | -0.488 | 0.065 | Ambiguous |
| 10 | [T10] | -0.369 | 0.176 | Ambiguous |
| 11 | [T11] | -0.426 | 0.113 | Ambiguous |
| 12 | [T12] | 0.009 | 0.975 | Ambiguous |
| 13 | [T13] | 0.189 | 0.501 | Ambiguous |
| 14 | [T14] | -0.672 | 0.006 | Cold |
| 15 | [T15] | 0.783 | 0.001 | Hot |
| 16 | [T16] | -0.592 | 0.02 | Cold |
| 17 | [T17] | 0.158 | 0.574 | Ambiguous |
| 18 | [T18] | -0.383 | 0.159 | Ambiguous |
| 19 | [T19] | 0.906 | 0.0 | Hot |
| 20 | [T20] | -0.707 | 0.003 | cold |

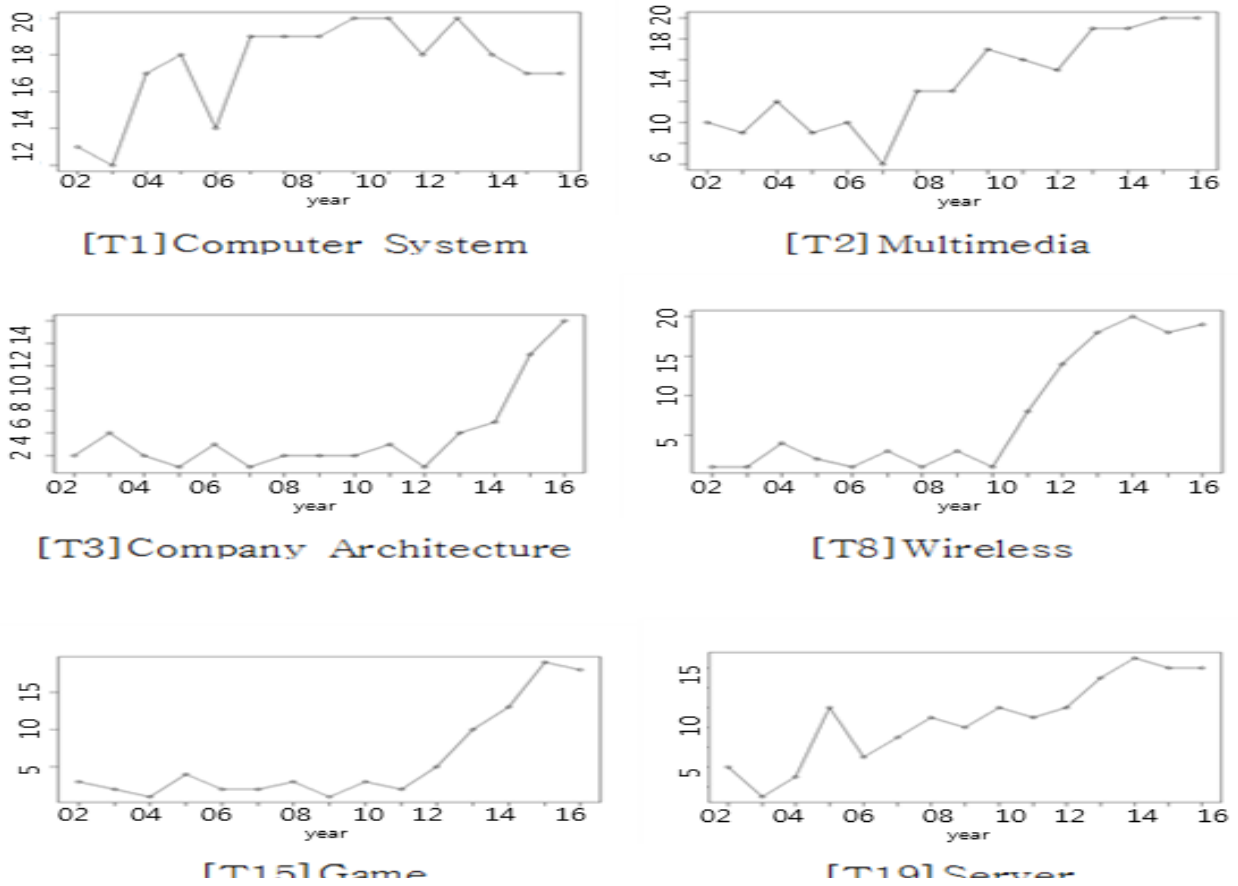


Fig.6 Six hot AI technologies 1

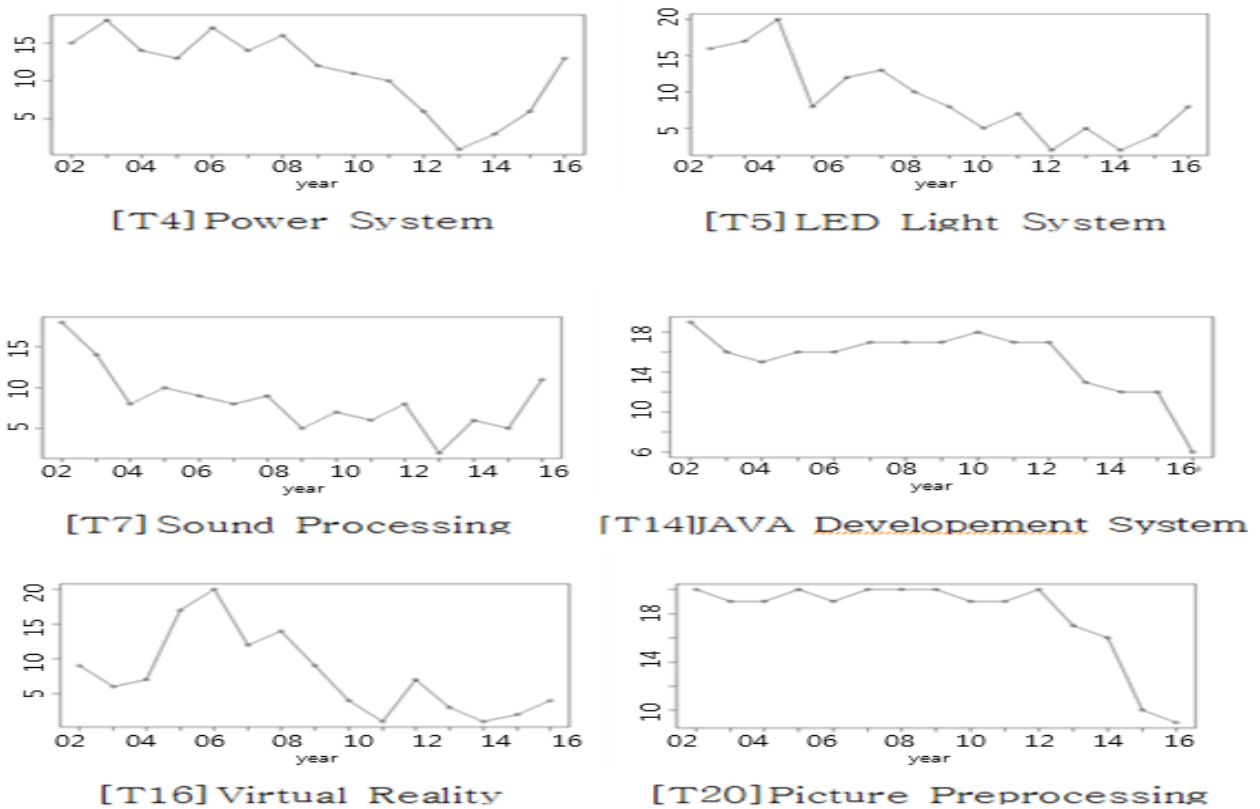


Fig.7 Six cold AI technologies 1

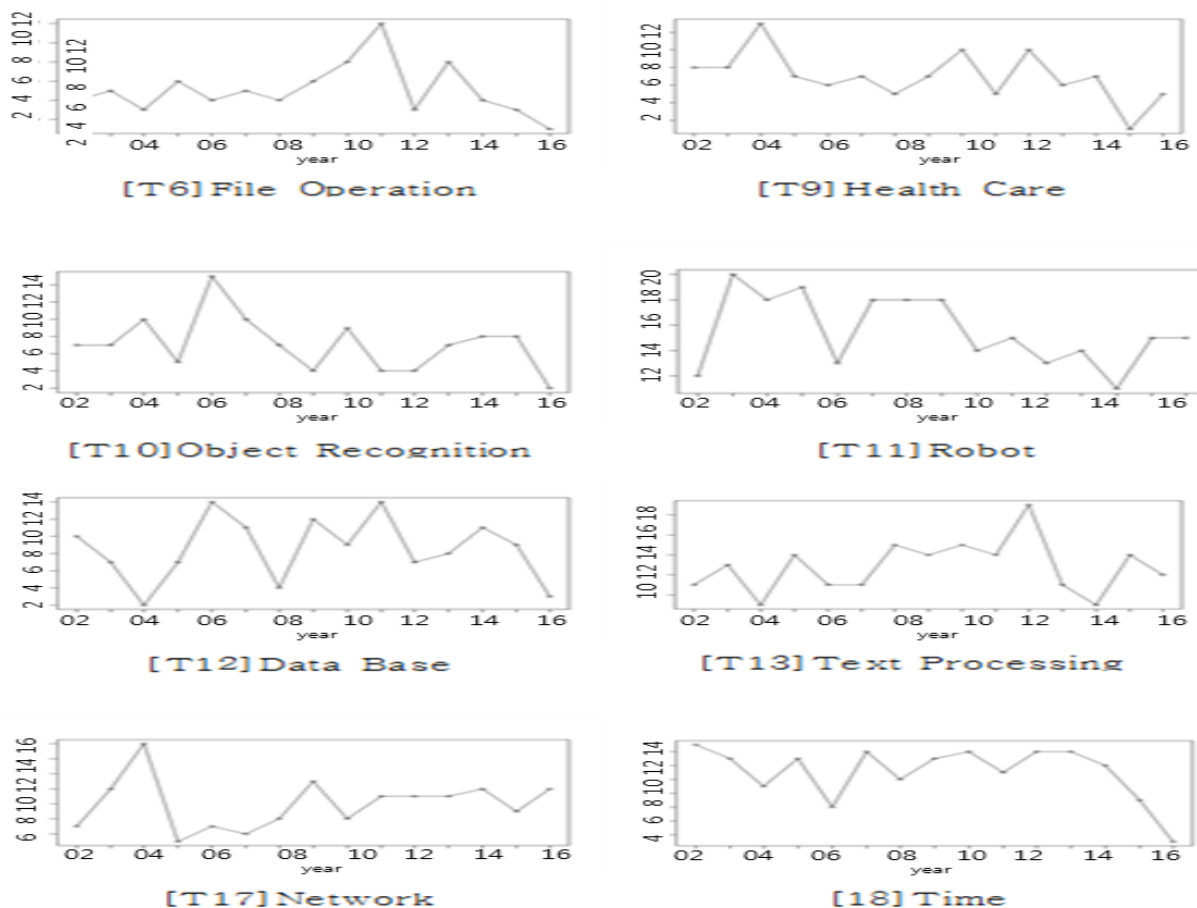


Fig.8 Eight ambiguous AI technologies 1

4.4 Verification

Six hot technologies and six cold technologies were derived from the trend analysis of the AI topics. The six hot technologies were [T1] computer systems, [T2] multimedia, [T3] company architecture, [T8] wireless technology, [T15] games, and [T19] servers. The six cold technologies were [T4] power systems, [T5] LED light systems, [T7] sound processing, [T14] JAVA development systems, [T16] virtual reality, and [T20] picture preprocessing.

The validity of the forecasted 12 technologies was verified using the recent 3-year (2014 to 2016) trends in the frequency of related words within each topic. The verification was based on the rate of matching between the forecast from the 15-year trend analysis and the forecast from the 3-year frequencies of related words within each topic. If the frequency of the related words in a topic recently increased in the abstracts regarding AI patents, then that topic was considered a hot technology. If the frequency recently decreased in the abstracts, then that topic was considered a cold technology.

The results of the trend analysis are shown in Table 5. The [T20] picture preprocessing technology could not be verified. The analysis of the matching rates demonstrated that 8 of the 11 technologies matched. The matching rate varied slightly depending on the method of preprocessing and the selection of related words within a topic.

5 Conclusion

Forecasting promising technologies can provide important basic information to establish a country's R&D investment strategies. To date, the Delphi technique, which requires technology experts, has been used to forecast technology. Although the technique has advantages, such as the effective distribution of research resources and the collection of opinions, it also has many problems, such as increased time and economic costs due to the complexity of its procedures and the necessity to mobilize several experts [15]. In this regard, text mining is emerging as an alternative to the Delphi technique. The development of a forecasting engine that enables easy, fast, and low-cost technology

Table 5 Results of the verification 1

| | | T1 | T2 | T3 | T4 | T5 | T7 | T8 | T14 | T15 | T16 | T19 | T20 |
|----------------|------|------|-----|-----|-----|-----|------|-----|-----|-----|-----|-----|-----|
| Frequency | 2014 | 1416 | 679 | 711 | 714 | 363 | 1018 | 659 | 727 | 657 | 431 | 642 | 203 |
| | 2015 | 1472 | 917 | 811 | 659 | 503 | 1094 | 726 | 671 | 800 | 469 | 714 | 285 |
| | 2016 | 1511 | 923 | 822 | 647 | 582 | 1107 | 801 | 653 | 894 | 689 | 759 | 205 |
| Trend Analysis | | ↑ | ↑ | ↑ | ↓ | ↓ | ↓ | ↑ | ↓ | ↑ | ↓ | ↑ | ↓ |
| Verification | | ↑ | ↑ | ↑ | ↓ | ↑ | ↑ | ↑ | ↓ | ↑ | ↑ | ↑ | Δ |
| Match | | O | O | O | O | X | X | O | O | O | X | O | Δ |

forecasting will facilitate future studies on forecasting in various fields.

In the present study, a science and technology trend forecasting engine was developed using the LDA topic modeling technique. First, a framework for the forecasting system was created to design the engine. Second, the workflows of the LDA topic modeling system and the technology trend analysis system, which were the components of the framework, were explained. Finally, the system's structure was defined in detail and designed using the class and sequence diagrams. The trend analysis of the AI topics resulted in six hot technologies and six cold technologies. The six hot technologies were computer systems, multimedia, company architecture, wireless technology, games, and servers. The six cold technologies were power systems, LED light systems, sound processing, JAVA development systems, virtual reality, and picture preprocessing. To verify the validity of the 12 forecasted technologies, the recent 3-year (2014 to 2016) changes in the frequency of related words within each topic were analyzed. In other words, the forecast from the 15-year trend analysis and the forecast from the 3-year frequencies of related words within each topic were compared in terms of their matching rates. As a result, eight of the 11 technologies matched (1 technology could not be verified). The present study on the development of the science and technology trend forecasting system using a text mining technique is likely to have academic and practice implications.

The practical implications of this study are as follows: First, the forecasting system that was developed using the text mining technique laid the foundation for easy, fast, low-cost, and automated forecasting. Second, file handling programs were developed to effectively process big data during

preprocessing and the analysis of technology trends. In addition, the study may have positive academic implications for future research in terms of reviewing various reference materials and building a theoretical basis for the design of forecasting engines.

The present study's limitations and future research directions are as follows: First, the Delphi technique is frequently used for technology forecasting, and thus future studies are required to identify whether this technique can also be applied to the developed forecasting system. Second, the coefficients for regression analysis were used to assess the trends of individual technologies in the technology trend analysis. Future research should be conducted to examine effective statistical methods other than the regression analysis. Third, this study analyzes changes in the frequency of related words within the AI topics to verify the forecasting system, and further research is necessary to identify more effective verification methods. Fourth, while the present study's scope extends to the verification of the forecasting system, follow-up studies may implement actual systems.

Acknowledgments: This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2015S1A3A2046781)

References:

- [1] Cho, B. S., Ji, K. Y., Kim, Y. J. and Lee, B. G., Future Technology Forecast Process, ETRI Planning Report, 2009.
- [2] Park, J. H. and Song, M., A Study on the Research Trends in Library & Information Science

- in Korea using Topic Modeling, Korea Society for Information Management, Vol. 30, No. 1, 2013, pp. 7-32.
- [3] Jern, S. H., Park, S. S. and Jang, D. S., Patent Analysis and Technology Forecasting, Kyowoosa, 2014.
- [4] 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.
- [5] Lee, S. K., Kim, S. I., Choi, C. T., Ahn, J. H., You, J. W. and Jerng, S. H., A Study on the Selection of the 10 Most Promising Technologies of KISTEP in 2016, KISTEP Research Report 2016-080, 2016.
- [6] 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, 2006, pp. 69-82.
- [7] Blei, D. M., Ng, A. Y. and Jordan, M. I., Latent Dirichlet Allocation, The Journal of Machine Learning Research, Vol. 3, 2003, pp. 993-1022.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] Jeong, B. K., and Lee, H. Y., "Research Topics in Industrial Engineering 2001~2015", Journal of the Korean Institute of Industrial Engineers, Vol. 42, No. 6, pp. 421-431, 2016.
- [12] 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.
- [13] 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.
- [14] Song, M. and Kim, S. Y., Detecting the Knowledge Structure of Bioinformatics by Mining Full-text Collections, Scientometrics, Vol. 96, No. 1, 2013, pp. 183-201.
- [15] Ko, B. Y. and Lo, H. S., Discovery of Promising Business Items by Technology-industry Concordance and Keyword Co-occurrence Analysis of US Patents, Journal of Korea Technology Innovation Society, Vol. 8, No. 2, 2005, pp. 860-885.