공학박사학위논문

A Public-based Exploratory Approach to Technology Foresight: Text Mining and Scenario Planning

대중 데이터를 활용한 탐색적 기술 예측: 텍스트 분석과 시나리오 기획

2020 년 2 월

서울대학교 대학원
산업공학과

김 영 준

Abstract

A Public-based Exploratory Approach to Technology Foresight: Text mining and Scenario Planning

Youngjun Kim

Department of Industrial Engineering

The Graduate School

Seoul National University

Today, the need for public opinion in the technology foresight process continues to increase. The perceptions of experts and public about the social impacts and risks that can arise from the proliferation of emerging technologies could differ. Meanwhile, due to the recent rapid development of information and communication, a large amount of information and knowledge, which were difficult to access in the past, are being shared online. Web-based data including people's opinion and prediction on various technologies are being accumulated exponentially. These changes provide the opportunity for new technology foresight activity beyond the qualitative approach that has traditionally been done for future research. In response, this study proposes a new framework for public-based exploratory technology foresight that analyzes more diverse future possibilities, overcoming limitations of normative technology forecasting. To this end, the following three objectives are covered through each research theme: (1) To suggest criteria for evaluating and screening data sources for public-based technology foresight (2) To present a research framework that identify trends in public issues about new technologies (3) To develop a technology foresight process that derive future scenarios from big data

The first research theme proposes a framework for selecting and validating data sources for public-based technology foresight. In other words, it finds out which of the many online communities have valuable data source. Specifically, we select various candidate communities where discussions and predictions related to technology are made and evaluated in terms of diversity and expertise.

The second research theme aims to analyze how the public accepts new technologies and how their perceptions change. By understanding and categorizing this process, it would be possible to infer how the public's perception of new technologies will change and respond in the future. To this end, we quantitatively analyze how the themes contained within text data change over time by utilizing time-series topic modeling technique.

The third research theme aims to identify the relationships among the factors related to the future of emerging technologies and to develop a scenario-based technology roadmap by analyzing big data. To this end, scenario planning is conducted based on fuzzy cognitive map, which is a directed graph that represents causal relationships between concepts. As a result, quantitative predictions of the future of emerging technologies are derived from data of non-experts, suggesting the potential for public-based foresight activity.

In whole, this thesis presents a new technology foresight and exploration framework for today's big data era through a sequential processes of screening, understanding, and interpreting online community data. To analyze unstructured text data, scenario planning utilizing keyword-based text mining techniques is conducted, and it is expected to be able to derive more exploratory and quantitative results. The public-based technology foresight framework presented in this study, which can identify various possibilities for the future of technology, is anticipated to have a complementary influence on conventional expert-oriented normative forecasting activities.

Keywords: Public-based technology foresight, Online community data, Text analysis, Topic modeling, Scenario planning

Student Number: 2014-21814

Contents

Abstract **i**

Contents **iv**

List of Tables **viii**

List of Figures **xi**

Chapter 1 Introduction **1**

1.1 Background and Motivation ····· 1

1.2 Research Objectives ····· 4

1.3 Scope and Framework ····· 6

1.4 Organization of the Thesis ····· 8

Chapter 2 Background **10**

2.1 Theoretical Background ····· 10

2.1.1 Public-based Technology Foresight and Big Data ····· 10

2.1.2 Online Communities for Technology Foresight ····· 12

2.2 Methodological Background ····· 15

2.2.1 Topic Modeling ····· 15

2.2.2 Fuzzy Cognitive Map-based Scenarios ····· 17

Chapter 3 Evaluation of Public Foresight Communities **20**

3.1 Introduction ····· 20

3.2 Backgrounds ····· 22

3.2.1 Assessment of Online Communities ····· 22

3.2.2 Expertise and Diversity of Online Community ····· 23

3.3 Measurement of Expertise and Diversity ····· 25

3.2.1 Expertise ····· 25

3.2.2 Diversity ····· 26

3.4 Illustrative Case Study ····· 28

3.4.1 Data Description ····· 28
3.4.2 Data Analysis ····· 29
3.4.3 Measuring Expertise and Diversity ····· 32
3.5 Conclusion ····· 37

Chapter 4 Public Understanding of Technology 39
4.1 Introduction ····· 39
4.2 Backgrounds ····· 42
4.2.1 Understanding Public Interest ····· 42
4.2.2 Public Attitudes on Emerging Technologies ····· 43
4.3 Proposed Approach ····· 45
4.4 Illustrative Case Study ····· 47
4.4.1 Data Collection ····· 47
4.4.2 Topic Extraction ····· 48
4.4.3 Identifying Relationship Between Topics ····· 50
4.4.4 Finding Pathways of Changing Interests ····· 52
4.5 Conclusion ····· 61

Chapter 5 Scenario-based Foresight 63
5.1 Introduction ····· 64
5.2 Scenario-based TRM ····· 66
5.3 Proposed Approach ····· 68
5.3.1 Identification of Internal and External Factors ····· 68
5.3.2 Cause-effect Relationship Analysis ····· 69
5.3.3 Development of Weighted TRM ····· 70
5.3.4 Scenario Development ····· 71
5.3.5 Dynamic Simulation of Scenarios ····· 72
5.3.6 Multiple Scenario Roadmapping ····· 72
5.4 Illustrative Case Study: Unmanned Aerial Vehicle (UAV) roadmap ····· 74
5.4.1 Data Collection ····· 74
5.4.2 Identifying Internal and External Factors ····· 75
5.4.3 Cause-effect Relationship Analysis ····· 78
5.4.4 Development of Weighted TRM ····· 82
5.4.5 Scenario Development ····· 83
5.4.6 Dynamic Simulation of Scenarios ····· 84
5.4.7 Final Scenario Roadmapping ····· 86

5.5 Conclusion ····· 89

Chapter 6 Concluding Remarks **91**

Bibliography **97**

Appendix **109**

국문초록 **116**

List of Tables

Table 2.1 Types and examples of online communities for technology foresight ····· 13
Table 2.2 Comparison of FCM-based scenarios with Trend Impact Analysis ····· 19

Table 3.1 Measurement of expertise and diversity of online community ····· 27
Table 3.2 Result of data collection ····· 28
Table 3.3 Data analysis result ····· 31
Table 3.4 Diversity and expertise measurement result ····· 33
Table 3.5 Description of selected communities ····· 36

Table 4.1 Summary of topic extraction ····· 51

Table 5.1 Data sources for UAV technology ····· 75
Table 5.2 Results of concepts extracted by topic modeling ····· 77
Table 5.3 Classification of concepts and their probabilities ····· 78
Table 5.4 Result of association rules ····· 79
Table 5.5 Result of causal matrix ····· 81
Table 5.6 Morphological matrix for scenario development ····· 84
Table 5.7 Result of dynamic simulation ····· 85

Table 6.1 Theoretical and practical contribution of this thesis ····· 92

List of Figures

Figure 1.1 Overall structure of this thesis ····· 9

Figure 2.1 SVD method of LSA ····· 16

Figure 3.1 Diversity-expertise map ····· 36

Figure 4.1 Overall framework ····· 46
Figure 4.2 An example of finding a pathway of interest ····· 47
Figure 4.3 Examples of articles of Futuretimeline ····· 49
Figure 4.4 An example of finding optimal number of topics ····· 50
Figure 4.5 Overall linkages of topics ····· 52
Figure 4.6 VR topic pathway ····· 55
Figure 4.7 AI topic pathway ····· 56
Figure 4.8 Space topic pathway ····· 57
Figure 4.9 Medical topic pathway ····· 58
Figure 4.10 Society topic pathway ····· 59
Figure 4.11 Environment topic pathway ····· 60
Figure 4.12 Rest topics ····· 61

Figure 5.1 Overall framework ····· 69
Figure 5.2 An example of causality matrix ····· 71
Figure 5.3 An example of a weighted TRM ····· 72
Figure 5.4 Scenario development based on combination of concepts ····· 73
Figure 5.5 An example of multiple scenario roadmaps ····· 74
Figure 5.6 Part of the generated term-document matrix ····· 76
Figure 5.7 Result of weighted TRM development ····· 83
Figure 5.8 Centrality analysis of weighted TRM ····· 83
Figure 5.9 Roadmap for scenario 1 ····· 87

Figure 5.10 Roadmap for scenario 2 ····· 87
Figure 5.11 Roadmap for scenario 3 ····· 88

Chapter 1

Introduction

1.1 Background and Motivation

Exploring the future of new technologies or innovations has always been one of the most important issues of future research. Various methodologies have been developed to predict how the emergence and development of new technologies will change human society in the future. Proactively capturing the impact and risks of new technologies is key to competitive advantage, not only at the company but also at the national level [60].

In line with this, technology forecasting has traditionally been a major topic for future research on technology. While numerous terms are used to describe technology forecasting, the key goal of technology forecasting is to assess future goals, needs, desires and missions related to technology [34]. In other words, analyzing what ought to be or needs to be possible at some future time is a decisive theme of technology forecasting. Thus, the results of these activities substantially aims to provide groundwork to allocate technology-generating resources such as investment, human resource and others in order to reach organizational objectives. Accordingly, the perspectives dealing with the future in technology forecasting tend to be deterministic, and the accuracy of prediction is a high priority [21, 26]. Representative methodologies utilized for technology forecasting include cross impact analysis [41], growth curves [74], extrapolation [6], and system dynamics [35].

However, recent rapid transitions in technological advances, represented by radical innovation, discontinuous innovation, and disruptive innovation, are changing the paradigm of future research [108]. First, as the technology life cycle shortens, the uncertainty of emerging technologies and innovations increases [4]. Next-generation technologies often

emerge even before the knowledge and data associated with emerging technologies is sufficiently accumulated. In addition, the convergence of interdisciplinary technologies has deepened the influence among non-technological factors such as society, economy, politics, and environment [23]. In the past, an approach to predict future changes by extrapolating the trends in the data associated with technology, namely forecasting, was relatively straightforward. On the other hand, today it is almost impractical to predict the exact future through projection.

In consequence, the orientation of future research on technology is shifting from a normative approach, which sets out the desired future we want in advance and studies what to do to reach that goal, to exploratory foresight activities [90]. In other words, the emphasis is now on the technology foresight, which explores the various possible future changes that can occur as technology advances rather than aiming to predict one accurate future [29]. The process of creating a shared vision of the future can facilitate stakeholder collaboration and bring about a shared sense of ownership to the implementation strategy. As a result, efforts are being made to create long-term and opportunities for society through science and technology and innovation in a direction different from conventional technology forecasting.

In this context, the scope of technology foresight is broadening to include users and the public instead of considering the opinions of a few experts, a large number of users and communities are involved in predicting the future society. The background of this change is as follows. First, there are limitations of the conventional expert-oriented approach [80]. Expert-based forecasts may have a limited scope for their analysis or results. Since they make predictions and analyzes around their area of expertise, there is a limit to exploring possibilities for more areas [7]. As a result, expert-driven technology foresight may eventually overlook the potential application or impact of technology. Second, as information accessibility improves, the gap between experts and the public is narrowing and the boundaries between the two groups are blurring [15]. In addition, as the importance of collective intelligence and the importance of scientific communication is emphasized, the necessity of future research through participants from more diverse backgrounds is increasing. This paradigm shift opens up the possibility of foresight activity through public-

based technology foresight.

The following examples illustrate these recent changes. The most representative is foresight 6.0 [22], which jointly explores the future that can arise from big data in the era of the Fourth Industrial Revolution. The main emphasis of this concept is that the domain of key participants in forecasting activities has been extended to netizens. In particular, futurizen, which refers to people who contribute personally to future research related to their interests online, is emerging as a new subject of future research [39].

1.2 Research Objectives

Limitations of the existing public-based technology foresight researches to be overcome in this thesis can be classified theoretically and methodologically. From the theoretical perspective, there are few examples of quantitative and empirical analysis of public perceptions of technology. In other words, most studies analyzing how the public perceives new technologies and forecasts the future associated with them have mostly relied on indirect indicators or attempted qualitative approaches. As a result, it is possible that the results of the forecasting activities may not reflect the public opinion practically. From the methodological perspective, the limitation is that the data collection process depends heavily on methods such as surveys and interviews. In this case, the number of participants could be limited due to the time and cost constraints and it thereupon lead to less diversified alternatives than utilizing collective intelligence or big data. Thus, this thesis aims to address the following three objectives through each of the research topics.

(a) To suggest criteria for evaluating and screening data sources for public-based technology foresight
(b) To present a research framework that identify trends in public issues about new technologies
(c) To develop a technology foresight process that derive future scenarios from big data

The first research theme proposes a framework for selecting and validating data sources for public-based technology foresight. In other words, it finds out which of the many online communities have valuable data source. Specifically, we select various candidate communities where discussions and predictions related to technology are made and evaluated in terms of diversity and expertise. We present a quantitative framework for evaluating this and ultimately identify online communities that are considered to be of high value for use in technology foresight.

The second research theme aims to analyze how the public accepts new

technologies and how their perceptions change. By understanding and categorizing this process, it would be possible to infer how the public's perception of new technologies will change and respond in the future. To this end, we quantitatively analyze how the themes contained within text data change over time by utilizing time-series topic modeling technique.

The third research theme suggests a framework for deriving integrated foresight scenarios related to specific technologies from online communities. FCM (Fuzzy Cognitive Map) technique that can draw quantitative scenario through causality of various concepts is mainly utilized. In order to empirically apply predictions from the online community for actual technology planning and strategy, we build FCM based on technology roadmap and build final scenarios.

1.3 Scope and Framework

The entire scope of the thesis is to present a systematic public-based technology foresight framework through a stepwise approach to data evaluation, data interpretation, and data analysis. Thesis consists of three research themes for each step.

The first step is to determine which of the many online communities are suitable for public-based technology foresight. There are some studies that utilize data from the online community for future research, but few quantitatively assess and validate each data source in advance. In particular, evaluating online communities for specific purposes of technology foresight is important at the moment when non-experts are becoming increasingly important in future research. At this stage, we analyze several online communities where users share their opinions and predictions about various emerging technologies. As a result, online communities deemed valuable for technology foresight are evaluated in terms of diversity and expertise.

The second step aims to analyze and categorize changes of perception of public on future technologies from previously selected data to help interpretation. It is a process of analyzing and grasping the public's acceptance of emerging technologies before the comprehensive technology foresight process. There is a Hype Cycle as a representative existing model of the public expectations of technology, but there is a clear limitation that this model is made only by a qualitative approach. Thus, this research theme suggests a framework that examine the relationship between the public and emerging technologies through a quantitative and practical approach. To this end, we analyze how public perceptions change in relation to the various types of emerging technologies generated by technology-related online communities. In other words, it is a step that provides insight into how to utilize the data in the future technology foresight process by understanding the previously selected data.

The final step is to finally derive the technology foresight scenario from the public data. Namely, previously validated data are utilized to make predictions about specific

emerging technology. To this end, a framework is provided to empirically analyze how technology-related data generated in the online community can be used in the technology foresight process. This includes case study through illustrative example.

Through this sequential processes, this thesis covers the process of screening data for technology foresight, analyzing the relationship between the public and emerging technologies, and deriving scenarios for emerging technologies. In other words, the entire scope of the thesis is to present a systematic public-based technology foresight framework through a stepwise approach to data evaluation, data interpretation, and data analysis.

1.4 Organization of the Thesis

This thesis is composed of six chapters as shown in Figure 1.1. The remainder of this thesis is organized as follows. Chapter 2 introduces the theories and methodologies that are covered in the main research themes. Specifically, the theoretical background section discusses public-based technology foresight and related online communities through literature reviews. The methodological background section covers the text analysis techniques and fuzzy cognitive maps utilized in this thesis. Chapters 3, 4, 5 are main bodies of this thesis. Three research themes – public foresight communication evaluation, public understanding of technology, and scenario-based foresight – are covered in these chapters respectively. Finally, Chapter 6 summarizes the results and limitations of this thesis.

	Chapter 1. Introduction		
	Chapter 2. Background		
	Chapter 3. Public Foresight Communities Evaluation	Chapter 4. Public Understanding of Technology	Chapter 5. Scenario-based Foresight
Purpose	Evaluating appropriate **public data source** of future studies	Analyzing the **changes of public perception** on future technologies	Deriving future **scenarios** online communities quantitatively
Method	Bibliometrics Keyword analysis	Time-series Topic Modeling	Fuzzy Cognitive Maps Technology Roadmap
Scope	Various future technologies		One target technology
	Chapter 6. Conclusion		

Figure 1.1 Overall structure of this thesis

Chapter 2

Background

2.1 Theoretical Background

2.1.1 Public-based Technology Foresight and Big Data

Conventional and prevailing methods of gathering information on technology foresight include Delphi, workshops, brainstorming, interviews, or horizon scanning [47]. The commonality between these approaches is that they rely on information provided directly by a few experts. Information collected in this way can ensure high reliability and validity.

Yet, a concept that is considered crucial in today's technology foresight research is collective intelligence [50]. In other words, given the future of new technologies, exploration of more possibilities and opportunities should be based on diverse perspectives, knowledge and information. Naturally, the subject of technology foresight is changing from a few experts to a non-specialist, namely the public [40]. These trends are driving the development of more diverse methodologies and approaches for future research related to technology [1].

Data collection is paramount to the success of foresight activities that require the involvement of more diverse participants. Traditional methods of collecting information may limit the amount of data available due to economic and time constraints. The main alternative in this situation is the big data generated online. Today's online innovations, such as Web 3.0, provide an opportunity to overcome the limitations of direct knowledge gathering methods [8].

Big data has been one of the most emerging topics worldwide in recent years for a variety of industries and research. The term big data is defined in various ways, but it can be generally regarded as a data set with 3V characteristics of volume, velocity, and variety [76]. Due to these characteristics, big data requires a platform and techniques for analyzing huge amounts of data that is difficult to handle with conventional management and systems.

The reasons for the emergence of big data analysis can be summarized by market demand and supply factors. First, the amount and quality of information have grown rapidly through the smart revolution [97]. The smart revolution has popularized smart devices, resulting in a dramatic increase in data usage due to increased consumption of large-capacity content and increased use of social media. In addition, data output is rapidly increasing due to the popularization of the Internet of Things, including machine-to-machine sensors [43]. Second, the demand for data analysis has increased for enterprises seeking to increase their competitiveness by generating information from quantitative and qualitative data [25]. As market competition increases, the utilization of existing differentiation is reduced, and companies are increasingly seeking new competitive advantage by utilizing customers' consumption patterns and needs inherent in the big data. Third, the demand for big data analysis increased due to the combination of the development of computing technology and the necessity of creating a new market due to the saturation of existing ICT product market [86].

As such, opinions and predictions about a diverse group of people with more varied backgrounds are available on the web, and a circumspect and thorough analysis of them has become a major challenge. In other words, in today's big data era, public data in the online community presents the challenges of future research into a new paradigm.

Big data analysis generally follows the process of data generation, collection, storage, analysis, and expression [9]. There are various detailed areas and related techniques for each process. Especially, analysis techniques include text mining, classification analysis, association analysis, and cluster analysis. Text mining is a technique for finding hidden meaning and finding meaning from large amounts of text data. Since most of the big data on the Internet is in the form of unstructured data, a preprocessing is required to find

meaningful patterns in a large amount of unstructured text. Database structuring and mathematical algorithm are applied to the preprocessed text. Various preprocessing and analysis algorithms such as keyword-document matrix, topic modelling, and opinion mining are applied according to the purpose of analysis, data type, and research environment, and they are continuously being developed in terms of both methodology and application.

2.1.2 Online Communities for Technology Foresight

Over the past decades, the widespread popularity of online communities has been vastly provoked by the rapid development of information and communications technology. The paradigm of online communities has significantly shifted from the traditional Web to an interactive, participatory, and engaging Web, which is called Web 2.0 [81]. As a result, several hundred million users now participate in hundreds online communities every day. Zeng [112] showed that the quality of contents from online communities can exceed the quality achieved from traditional sources of knowledge such as information specialists, and librarians.

These online communities consist of group members based upon shared enthusiasm for an issue or an activity. However, each group possesses different attributes and shows a wide variety of characteristics. Several researches provide different classification schemata for online communities and its attributes [65]. They can assist in identifying online communities that are similar. Whittaker et al. [107] classified online communities by attributes such as a shared goal or interest, intense interactions and strong emotional ties, shared activities between community members, access to shared resources, and social conventions. Hagel and Armstrong [45], broadly partitioned the space of online communities into four areas: Communities of interest, communities of relationship, communities of fantasy, and communities of transaction. Also, Lechner et al. [66] has defined the types of community as discussion or conversation communities, task and goal-oriented communities, virtual worlds and hybrid communities. Plant [84] proposed three-dimensional taxonomy of the online community space which considers the degree of community regulation, the degree of community openness to membership and the degree to

which a community is involved in for-profit activities.

Building upon the previous research, online communities that conventionally utilized for technology forecasting can be categorized as shown in Table 2.1 [93]. This study aims at technology foresight that explores a wider range of future possibilities, rather than technology forecasting which is purposed to predict future accurately. Therefore, we do not follow the existing taxonomy of online communities that were mainly used for forecasting. Most of these websites are expert-oriented and the attitude for the future is normative rather than exploratory. In response, the discovery of new data sources for technology foresight is the starting point of this thesis. The types of online communities utilized in this study mainly belong to news and blogs that are familiar to ordinary people and users. An analysis is conducted in subsequent chapters to identify which online community provides valuable information.

Table 2.1 Types and examples of online communities for technology foresight

Types	Descriptions	Examples
Databases and Wikis	• A digital archive that provides a classification schema for future-related information • Used for mapping strategic foresight and horizon scanning	LongBets, iKnow, Trendradar2020, Shaping Tomorrow, SigmaScan, Delta Scan, TechCast, Forecasting World Events, The Seven Horizons, wrong tomorrow, Future Scanner, TechCast, TrendWiki
News and Blogs	• Including replies that offer technological trends and predictions of the individual and their interaction in online websites	Gartner, McKinsey, MIT Technology Review, Kurzweil Accelerating Intelligence, Next Big Future, World Future Society, Long Bets, iKnow, SigmaScan, TechCast, TechCrunch, TheVerge
Social Rating Systems	• The vest set of assumptions, predictions, conjectures, and the rates of them	Is it Future proof?, bean sight, The Future of Facebook Project, predicto.net, wefutr, Web of Fate, Wikistrat, The Future of Facebook

	•	Using the scales like relevance, impact, likelihood or desirability	Project, Forecasting ACE, NY Times Technology Timeline
Collaborative Scenarios	•	Focus on pooling only potential scenarios and solutions to specific future challenges and the ability to aggregate assumptions about the future into scenarios	superstruct, signtific map, Risk Interconnection Map
Prediction Markets	•	The payoff of contracts depends on the possible events in the future	intrade, inklingmarket, iPredict, Popular Science Predictions

2.2 Methodological Background

2.2.1 Topic Modeling

Due to the advancement of information and communication technology today, a lot of data is coming out from various media such as news, blog, social media, and it needs a tool to handle it. One method is topic modeling, which aims to extract topics from a large set of unstructured documents [11]. By modeling word usage patterns and linking documents representing similar patterns, topic modeling has emerged as a powerful new technology for finding useful structures in unstructured document sets. Because of these characteristics, topic modeling is being used to analyze documents in various fields; for example, topic modeling has been used in computer vision [72], social network [52] and bioinformatics [70].

One of the earliest algorithms of topic modeling and still in use today is Latent Semantic Analysis (LSA), developed in the late 20th century [31]. In the LSA technique, it is assumed that words with similar semantics appear in similar parts of the text. Therefore, in order to find the semantic textual pattern of these similar words, Singular Value Decomposition (SVD) is performed to divide the term-document matrix (TDM) into low-dimensional matrix products. Figure 2.1 shows the SVD method for TDM. U is the impact of each word $t_1, t_2, t_3, \dots, t_n$ on the semantic textual patterns $v_1, v_2, v_3, \dots, v_p$, Σ is the importance of each semantic textual pattern, V means the effect of each document $d_1, d_2, d_3, \dots, d_m$on the semantic textual pattern. In this study, semantic textual patterns resulting from LSA are considered as keyword clusters, and each keyword cluster is used as a concept node of FCM.

The most popular topic modeling algorithm is Latent Dirichlet Allocation (LDA). LDA is a probabilistic model of what topics are present in each document for a set of documents or individual documents [12]. LDA analysis yields the set of words as many as the number of predefined topics and the probability of an individual word in a set. The topic is not derived from analysis, but is determined by the analyst by looking at the words in the

word set. Furthermore, it is possible to determine what content the document contains, depending on which topic is frequently included in the analysis document.

Meanwhile, topics of document and words that make up the topic may change over time. For example, even in the same field, important topics, data, methodology, or implications are dependent on time. There are several methodologies for capturing changes in these topics. Blei & Lafferty [11] analyzed the topic changes in science through dynamic topic modeling and found that prediction performance of dynamic topic modeling was better than static topic modeling. Wang et al. [105] proposed continuous time dynamic topic modeling as a way to overcome the memory requirement and computational complexity of time discretization.

Methods for analyzing the emerge, disappear, and change over time of these topics have been used in many areas. Zhang et al. [116] identified scientific topics over time and visualized their relationships. Chen et al. [18] analyzed the underlying topic of patent claims over time using LDA and predicted future trends. Kim & Oh [57] proposed a framework that uses LDA to analyze important topics and issues in Web news. Building upon the previous research, this study applies the LDA to large unstructured documents to extract topics that represent interest of the public, and identify changes of interest over time.

$$TDM = U\Sigma V^T$$

$$U = \begin{array}{c|cccc|} & v_1 & v_2 & \cdots & v_p \\ t_1 & u_{11} & u_{12} & \cdots & u_{1p} \\ t_2 & u_{21} & u_{22} & \cdots & u_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ t_n & u_{n1} & u_{n2} & \cdots & u_{np} \end{array} \quad \Sigma = \begin{array}{c|cccc|} & v_1 & v_2 & \cdots & v_p \\ v_1 & \Sigma_1 & 0 & \cdots & 0 \\ v_2 & 0 & \Sigma_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ v_p & 0 & 0 & \cdots & \Sigma_p \end{array} \quad V = \begin{array}{c|cccc|} & v_1 & v_2 & \cdots & v_p \\ d_1 & v_{11} & v_{12} & \cdots & v_{1p} \\ d_2 & v_{21} & v_{22} & \cdots & v_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ d_m & v_{m1} & v_{m2} & \cdots & v_{mp} \end{array}$$

Figure 2.1 SVD method of LSA

2.2.2 Fuzzy Cognitive Map-based Scenarios

Fuzzy Cognitive Map(FCM), which means a fuzzy graph structure expressing causal relationships, analyses the causal relationships among various concepts in order to extract expert knowledge [62]. Namely, FCM is a method of quantifying and expressing the causal relationship between concepts that are difficult or impossible to quantify, using knowledge of experts. In the FCM, concept nodes are represented as fuzzy sets and causal relationships between concept nodes are explained by fuzzy relations. The components of fuzzy awareness are as follows [61].

- Concepts: $C_1, C_2, C_3, \dots, C_n$. Important factors that can be driver or constraint on the topic issue under consideration.
- State vector: Expresses the concept value between 0 and 1 and A_i is a value indicating the state of the concept node A_i. The dynamic change of the state vector is the main result of FCM.
- Direct edges: $C_i \rightarrow C_j$. Refers to the relationship between concept nodes, and is represented by an arrow in the direction graph.
- Adjacency matrix: $E = (W_{ij})$. W_{ij} is the weight of the direct edge corresponding to $C_i \rightarrow C_j$, which has a value between -1 and 1, and the association matrix represents all relations between all concepts.

Recently, FCM has been used in scenario analysis and planning. Scenario planning has been used extensively in technology planning and strategic analysis because it can improve the ability to respond to uncertainty and the usefulness of the entire decision-making process through scenarios. In particular, FCM-based scenario analysis is an integrated method of FCM model in scenario planning, generally following the following steps [3, 55].

① Scenario preparation: Determine goals, time, and bounds of the scenario project

② Knowledge capture: Identify relevant concepts and potential scenario drivers from experts and previous studies, and integrate mental models of various experts into a final conceptual FCM scenario model.

③ Scenario modeling: Connect causal relationships, measure causal weights, and set a threshold for all concepts.

④ Scenario development: Set various input vectors through plausible combinations of concept nodes. Various scenarios are generated depending on whether the concept's state vector is 0 (not occurring) or 1 (occurring).

⑤ Scenario selection and refinement: Evaluate and improve scenarios by simulation according to the input vectors determined above. The scenario simulation is performed by updating the state value of the concept node over time. When this interaction process is repeated, the state value fluctuates and stabilizes, and converges to a certain value. This value is used as the output vector of the first input vector.

⑥ Strategic decisions: Strategic decisions for the long-term future are made using developed scenarios.

FCM-based scenario approaches elicit diverse experts' knowledge of uncertain driving forces that shape the future and simulate what-if experiments to create the alternative scenarios. Table 2.2 shows a comparison between FCM and another representative scenario analysis technique, Trend Impact Analysis, in terms of purpose, characteristics, and data.

Table 2.2 Comparison of FCM-based scenarios with Trend Impact Analysis

	Trend Impact Analysis	**FCM-based scenarios**
Purpose	Designed primarily for the evaluation of one key decision or forecast variable, which is quantitative and have the historical information	To develop scenario consider considering all drivers and causal links in the FCM
Characteristic	This method does not evaluate possible impacts that the events may have on each other	Causal links highlights impact of concepts on each other
Data	Process is sometimes constrained due to unavailability of reliable historic time series data	Historic time series data of the important trends is not required

Chapter 3

Evaluation of Public Foresight Communities

3.1 Introduction

In traditional technology foresight processes, experts play an important role and are the basis for the use of many methods. Their expert knowledge generates harmonized descriptions about possible future directions [93]. These traditional closed-loop foresight activities have lead experts to foresight with various cognitive limitations.

In line with this, technology foresight is now acknowledged not only by the area of experts, but also a discipline of general public [75]. Instead of simply relying on experts discussing future developments, new approaches also include external sources, such as suppliers, research institutes, users, online communities. By integrating such external sources, the potential of different points of view can be integrated into the foresight process, resulting in collective intelligence [77]. Further, a large number of users and communities are involved in predicting the future society, and their influence has been increased owing to reduced information gap and better accessibility [108]. Combining the concepts of corporate foresight with the research on open and user innovation leads to a recently developed process called “open foresight” [32].

Besides, the technological development of the Internet in the direction of a participative approach, the so called Web 2.0 [81], changed usage behavior dramatically. This evolution is characterized by user-centered and interactive websites and forums which foster user activities such as co-creation and communication, as well as content sharing and creation [54]. With these developments, online communities emerged and individuals are unified by shared interests or common goals and discuss these using an Internet platform on

the Web 2.0 [112]. Moreover, online communities possess the required expertise for foresight and might help by contributing to a more comprehensive understanding of the future. Their knowledge makes them especially valuable for foresight processes and therefore, a systematic integration of online communities might reduce uncertainty about future changes.

Yet, since a mass of websites exist in the Internet, it is necessary to identify the 'right' online communities [112]. To exploit the myriad of data from many online communities, it is essential to avoid the harm of garbage-in garbage-out. In particular, more careful data screening is needed to sort online communities for the specific purpose of technology foresight. However, most studies that utilized online communities' data as a source of information for technology foresight focused only on the external characteristics of the online communities such as purpose of community. Reliable public opinion mining requires that online communities be categorized based on actual intrinsic contents, activities and characteristics of online communities. Hence a systematic and quantitative evaluation and classification framework for the online community is now needed for public-based foresight process.

In response, this study proposes a new online community assessment framework for technology foresight. Specifically, it aims to quantitatively evaluate the elements that make up the online communities where technology-related discussions take place. In addition, the focus is on two criteria, expertise and diversity, to select technology-related posts from various users.

The remainder of this chapter is organized as follows. First, the related works section deals with a literature review of the characteristics of online communities and the techniques used to evaluate them. Next, the online community evaluation framework proposed in this study is explained in detail. Then, the results of applying this technique to the actual 17 technical online community candidates are illustrated. Finally, our discussion and concluding remarks are provided.

3.2 Backgrounds

3.2.1 Assessment of Online Communities

Due to the rapid development of information and communication technology, countless online websites have occurred since the late 20th century, and researches for utilizing them have been conducted in various fields. Especially in the 2010s, with the spread of smart phones, the limitation of space and time is almost eliminated, and the expansion speed of the sea of information is increasing exponentially. Under the circumstance, selecting an online community for a specific research purpose is now being an important task.

Pioneering studies in this regard have been made mainly in the field of computer science. PageRank is an algorithm used by Google Search to rank websites [48]. It works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites. Since the development of the PageRank algorithm, various subsequent studies have been conducted. Expertise Network [113] was applied to produce user expertise ranking of a Java developer bulletin board. It measures expertise scores through the question-answer network. Also, SPEAR(Spamming Resistant Expertise Analysis and Ranking algorithm) [110] was developed to estimate the quality of shared items and the expertise of users with respect to a particular topic of interest at the same time. Likewise, ExpertRank [106], evaluates expertise based on both document-based relevance and one's authority in his or her knowledge community. It modified the PageRank algorithm to evaluate one's authority so that it reduces the effect of certain biasing communication behavior in online communities.

As such, evaluation algorithms for websites have been continuously developed. Yet, most are not evaluations of the community itself, but rather the evaluation of individual users or specific content. Conventional algorithms mainly aim to assess the impact of a user or to measure the linkage of a web page. Thus, there may be limitations to selecting the community, which can be a pool of data sources.

3.2.2 Expertise and Diversity of Online Community

To develop a framework for evaluating online communities, we first need to discern their characteristics. Since this study aims to screen online communities that can be a data source for public-based technology foresight, literature review was conducted on characteristics that are considered to be able to better reflect technology-related predictions of people rather than the general characteristics of online communities.

There are several characteristics on which the online community can be useful in foresight activities. First, the expert knowledge available from users in various domains is essential to foresight activities [112]. It is clear that the main reason why participants in future studies could be extended from a small number of experts to a large number of users is that the expertise of the public has increased due to the narrowing of the information gap. In addition, it is notable that there is no time and space limitation in the communication between users in the online community. This boosts intense interaction between users and provides additional data generation. Since most of the discussions and articles in online communities are free of charge and accessible, this is an advantage for researchers. Further beneficial is the size of online communities. The number of users in each online community can range from as few as hundreds to as many as tens of millions. Correspondingly, this scale guarantees a thematically broad spectrum of the perspective of exploring the future [24].

In light of this, this study aims to establish two criteria of expertise and diversity in evaluating online communities. First, expertise in this context refers to the skills and knowledge of individuals in a particular technology [85]. Expertise is important for foresight because the factors affecting the future are manifold and complex and thus lead to a high degree of uncertainty. To reduce this uncertainty, sound knowledge from different areas is needed. In terms of online communities, content produced by highly-expertized users can make the community more active and often become lead users in the long run [10]. Second, diversity is an indicator of how varied the themes and scope of the technologies being addressed in the community are. The complexity of today's problems requires an inclusion

of different perspectives. Individual experts often have limited capability to solve complex and interdisciplinary problems alone and are therefore dependent on interactions [92]. Online communities offer a mixture of different area promoting features of the Web 2.0, and their diversified themes could be a promising way for diversity in foresight activities.

3.3 Measurement of Expertise and Diversity

This section describes the online community evaluation framework for technology foresight presented in this study and shows the results of applying it to the actual online communities. Unmeasurable or external characteristics of the online community are not taken into account. Namely, demographic information such as gender and age of users, and accessibility and structure of community are not considered in metric. Instead, a bibliometric analysis and text analysis of the data are performed, focusing on the data itself to be utilized in the actual foresight activity. Table 3.1 provides a summary of the ways in which online community diversity and expertise can be measured.

3.3.1 Expertise

Expertise of online communities can be measured with several indicators. First, members' profiles are used, providing information such as professional background, occupation and hobbies. These are certainly intuitive and direct information and can confirm the user's expertise. However, in most communities it is illegal or impossible to collect such personal information, which is not considered in the metrics proposed in this study.

User expertise also can be distinguished in user experience and technology-related knowledge [16]. User experience is measured by the total amount of post. However, measuring expertise by simply counting the number of posts or articles generated within a community can be rather one-dimensional and a leap of logic can exist. Whereas, user experience can be confirmed by the activity of users in the community [54]. To this end, members in the community are classified into three groups: Innovator/activist, tourist/crowd-follower, and lurker. If the ratio of innovator/activist who has the highest level of activity, that is, functioning as the lead user, is high, the user experience can be indirectly determined.

Besides, technology-related knowledge is measured by the amount of technical terms used in articles. In other words, by measuring how many professional or technological keywords occurred in the articles written by users, the expertise of the community is

determined.

3.3.2 Diversity

A variety of approaches are available to measure the diversity of content generated in online communities [112]. First, the number of members in community can be an indirect indicator. However, in this case, it may include users with no actual activity, and in some communities, it may be difficult to distinguish between users who post technological articles and those who do not.

Another approach is to analyze the average amount of posts per member. The average amount of posts per member is calculated by dividing the amount of posts by the number of members who have written a post in one thread. There is a good chance that more topics will exist in the community if many people post their opinions rather than having a small number of users writing a lot. However, it is difficult to apply this approach to a community where it is difficult to keep track of the number of users due to anonymity, or a community where a small number of people post and discuss within the post.

Diversity also can be measured by the number of themes, or topics, included in articles generated in a community. In other words, it is to analyze the contents of the articles written by users to find out how many topics are covered in a community. To this end, topic modeling techniques, which analyze latent topics in the document corpus through the probability generation model, can be utilized.

Table 3.1 Measurement of expertise and diversity of online community

	Criteria	Quantitative /Qualitative	Measure
Expertise	Knowledge level of keywords	Both quantitative and qualitative	Extracting keywords with high frequency and scoring each keyword with Likert scale
	Level of activity	Quantitative	Calculating ratio of innovator/active user: Innovators are considered as users who have written a lot of high impact articles
	Type of community	Qualitative	
	Profiles of authors	Qualitative	
Diversity	Ratio of author per document	Quantitative	Dividing the total number of authors by the total number of documents
	Optimal number of topics	Quantitative	Finding the optimal number of topics by topic modeling technique HDP

3.4 Illustrative Case Study

3.4.1 Data Description

To apply the online community evaluation framework suggested in this study to actual case, we investigated various websites. Since the purpose is to select an online community for technology foresight, we identified communities that provide opinions or information related to the future of emerging technologies. In order to gather information from users of various backgrounds, we targeted diverse types of online communities such as blogs, news sites, online forums, and social media.

20 online communities were selected as candidates as shown in Table 3.2. Then, we have crawled data related to emerging technologies or foresights from each website. Each website has a different format and type, so the process of extracting the data was also distinct. For example, news-driven communities, such as BusinessInsider, have a preliminary section of articles, so articles in technology categories were collected. In the case of FutureTimeline, on the other hand, since the purpose of the website itself is future research, the entire blog post was consistent with the purpose of this study. In the case of social media such as Twitter, retrieved results were collected through search queries related to technology. The description of the data collected from each source is shown in Table 3.2. As a result, 15396 articles were collected from a total of 20 communities.

Table 3.2 Result of data collection

Online Community	Number of articles	Number of authors	Description
BusinessInsider	780	54	Articles included in "TECH" category
Cnet	482	57	Articles included in technology-related topics such as 'Sci-tech', 'Smart homes', and 'Drones'
TechCrunch	822	71	Recent news articles regarding technology
Diamandis	301	30	Technology-related blog articles

Engadget	582	44	Articles included in "TOMORROW" category
FutureTimeline	1522	244	Blog articles on various technological fields
Twitter	632	167	Search results for keywords related to emerging technologies, such as "driverless", "smart home", and "quantum-computing"
io9	993	108	Recent blog articles regarding technology
KDnuggets	154	25	Articles regarding data science
Kurzweil AI	1932	121	Articles related to the latest technology trends
MIT Technology Review	364	72	Articles related to several emerging technologies
Quora	774	53	Articles included in "Technology Forecasting" and "Emerging Technology"
Reddit	954	279	Articles included in technology-related subreddits such as "technology" and "selfdrivingcars"
SingularityHub	1095	91	Recent news articles regarding technology
SingularityWebBlog	731	122	Blog articles on various technological fields
Techradar	196	21	Recent news articles regarding technology
TheVerge	723	88	News articles included in "TECH" category
Wired	1266	72	Articles related to several emerging technologies
World Future Society	822	144	Articles included in topics related to future technologies, such as "WorldFuture" and "Resources for Future-Minded Citizens"
ZDNet	271	71	Articles included in "Innovation" category

3.4.2 Data Analysis

Based on the text data collected from each community, preliminary work is performed to

measure diversity and expertise. This is done using a mixture of bibliometric approach and text analysis, the results of which are summarized in Table 3.3.

Metrics for calculating diversity are as follows. First, the ratio of author per document was used as an indicator of diversity. This is based on the assumption that the more people write, the more diverse the overall theme. As a result, the ratio of author per document was the highest in Twitter and ZDnet, and relatively low in BusinessInsider and Kurzweil AI.

In addition, the number of topics latent in the entire document corpus is also used as an indicator of diversity. It is based on the premise that the more various the topics covered in the community, the higher the diversity. The Hierarchical Dirichlet Process(HDP), a kind of topic modeling technique, is used to measure the optimal number of topics [103]. HDP is a nonparametric Bayesian approach to clustering grouped data. Unlike other topic modeling techniques, where the number of topics must be determined in advance, the HDP technique automatically derives the optimal number of topics. Therefore, it can be effectively used in the situation where the number of topics existing in the entire corpus should be checked. As a result of calculating the optimal number of topics, SingularityHub and Wired were the highest and Techradar was the lowest.

The following metrics are used to measure expertise. First, the knowledge level of the community is used as an indicator of expertise. In this study, knowledge level was calculated based on keyword and the process is as follows. High frequency keywords are extracted from the full text data collected by each community. The analyst then scores on the Likert scale for the top 50 keywords in frequency. Each keyword is qualitatively evaluated for how high it is related to emerging technologies. The average of these values is then evaluated to assess the knowledge level of each community. As a result of evaluating the knowledge level of each community, the World Future Society was the highest and the BusinessInsider and Quora were relatively low.

The level of activity of the community is also used to measure expertise. Level of activity can be assessed indirectly by the ratio of activist. In other words, the higher the percentage of activists, or lead users, the higher the expertise. Here, the percentage of

activists is calculated by identifying the users who write high impact articles for a topic. To this end, Latent Semantic Analysis(LSA), a technique that analyze relationships between a set of documents and the terms they contain by producing a set of concepts related to the documents and terms, is utilized [51]. Applying LSA can derive the impact of each document on each topic. In this study, activist is set up as a case where more than 50% of the articles written by each user are included in the top 20% of the topics with high impact. As a result of calculating the ratio of activist, MIT Technology Review was the highest and KDnuggets and BusinessInsider were relatively low.

Table 3.3 Result of data collection

Online Community	Degree of keyword expertise	Ratio of activist	Ratio of author per document	Optimal number of topics
BusinessInsider	1.36	0.15	0.07	8
Cnet	2.34	0.41	0.12	8
TechCrunch	4.36	0.38	0.09	16
Diamandis	3.48	0.24	0.10	15
Engadget	3.04	0.17	0.08	13
FutureTimeline	3.86	0.34	0.16	22
Twitter	2.0	0.28	0.26	18
io9	4.56	0.38	0.11	22
KDnuggets	2.68	0.14	0.16	12
Kurzweil AI	5.18	0.35	0.06	22
MIT Technology Review	4.82	0.44	0.20	19
Quora	1.08	0.17	0.07	17
Reddit	4.12	0.28	0.29	21
SingularityHub	4.3	0.29	0.08	25
SingularityWebBlog	4.54	0.33	0.17	18
Techradar	3.62	0.28	0.11	7
TheVerge	3.96	0.3	0.12	19
Wired	5.22	0.41	0.06	25

World Future Society	6.42	0.34	0.18	23
ZDNet	2.56	0.18	0.26	14

3.4.3 Measuring Expertise and Diversity

Based on the previously calculated results, the diversity and expertise values of each online community are finally derived. First, the values of level of knowledge, ratio of activist, ratio of author per document, and optimal number of topics are standardized. The sum of standardized level of knowledge and ratio of activist is the value of expertise, and the sum of standardized ratio of author per document and optimal number of topics is the value of diversity. The result of this calculation is shown in Table 3.4.

Table 3.4 Diversity and expertise measurement result

Online Community	Standardized ratio of author per document	Standardized optimal number of topics	Diversity	Standardized level of knowledge	Standardized ratio of activists	Expertise
BusinessInsider	-0.95	-1.66	**-2.61**	-1.70	-1.54	**-3.24**
Cnet	-0.26	-1.66	**-1.92**	-0.98	1.26	**0.28**
TechCrunch	-0.71	-0.22	**-0.92**	0.50	0.94	**1.44**
Diamandis	-0.52	-0.40	**-0.92**	-0.14	-0.57	**-0.72**
Engadget	-0.86	-0.76	**-1.62**	-0.47	-1.33	**-1.79**
FutureTimeline	0.33	0.87	**1.19**	0.13	0.51	**0.64**
Twitter	1.78	0.14	**1.92**	-1.20	-0.14	**-1.34**
io9	-0.39	0.87	**0.47**	0.65	0.94	**1.58**
KDnuggets	0.35	-0.94	**-0.59**	-0.73	-1.65	**-2.38**
Kurzweil AI	-1.04	0.87	**-0.17**	1.10	0.61	**1.72**
MIT Technology Review	0.85	0.33	**1.18**	0.84	1.59	**2.42**
Quora	-0.96	-0.04	**-0.99**	-1.90	-1.33	**-3.23**
Reddit	2.17	0.69	**2.86**	0.32	-0.14	**0.18**
SingularityHub	-0.75	1.41	**0.66**	0.46	-0.03	**0.42**
SingularityWebBlog	0.42	0.14	**0.56**	0.63	0.40	**1.03**
Techradar	-0.42	-1.84	**-2.26**	-0.04	-0.14	**-0.18**
TheVerge	-0.21	0.33	**0.11**	0.21	0.08	**0.28**
Wired	-1.12	1.41	**0.29**	1.13	1.26	**2.39**
World Future Society	0.53	1.05	**1.58**	2.01	0.51	**2.52**
ZDNet	1.75	-0.58	**1.17**	-0.82	-1.22	**-2.04**

In addition, the result of mapping the previous result by using diversity and expertise as two axes on the two-dimensional plane is shown in Figure 3.1. On this map, online communities with high diversity and expertise, which are located in the first quadrant, are finally selected as suitable source for technology foresight. A description of the 10 online communities that are considered advisable for public-based technology foresight is summarized in Table 3.4.

The online communities in the second quadrant have high expertise but low diversity. Therefore, online communities in this category may not be useful for exploring various future alternatives. However, they can be used to reflect expert opinions, not for public-based foresight. Blogs such as Kurzweil AI have a small number of participants who actually write, but each has a high level of professionalism. Thus, these communities can be considered as alternative data for existing expert-oriented foresight.

The online communities located in the fourth quadrant are considered to be more diverse but less expertized. As a result, online communities classified in this quadrant may be inappropriate for technology foresight that require more than a certain level of knowledge. However, they can be useful in identifying trends of public opinion or exploring various issues, rather than in-depth analysis of technology. In particular, in the case of Twitter, there are various studies using data generated from Twitter, and it is possible to analyze the network considering the connectivity of each Tweet.

Finally, the online communities in the third quadrant are rated both low in expertise and diversity. In other words, the community is considered to be difficult to use as a data source for technology foresight. However, there may be potential applications in the contents of the communities that are not captured by the evaluation criteria presented in this study. For example, Techradar and Engadget, located in the third quadrant, provide a large amount of detailed reviews of the latest electronics. Such review data may be used for analyzing functional features and technical characteristics of the product. While this may not serve the purpose for exploratory foresight, it is likely to be more useful than other categories of data in strategy planning.

The causes for some communities not finalized as suitable data sources are

considered to be as follows. First, Twitter has a large number of users and covers various topics. However, due to the regulation of the service, the length of each post is short and does not tend to include in-depth content. Therefore, Twitter's evaluation resulted in high diversity but low expertise. TechCrunch provides news and information related to technology, and the technological level of each article is high. Yet, as the community's focus is mainly on startups and businesses, it seems that various topics related to emerging technologies have not been derived. Lastly, in BusinessInsider, the weight of articles focused on companies or people related to technology was high. As a result, the focus on technology itself is relatively low, suggesting low expertise. In addition, as a relatively small number of journalists posted articles, diversity was also underestimated.

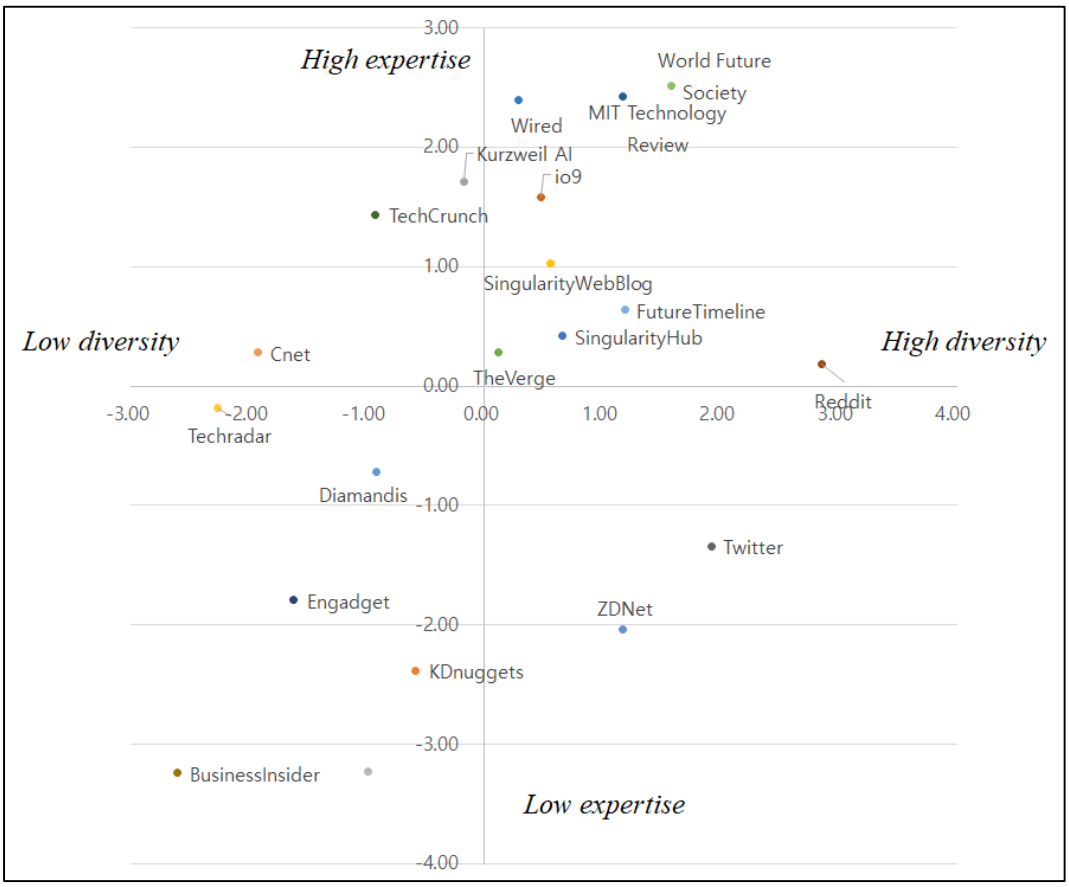

Figure 3.1 Diversity-expertise map

Table 3.5 Description of selected communities

Online Community	Description
TheVerge	• Representative news webzine specialized in science and technology • Articles are grouped by promising technology area • Each article contains user comments
SingularityHub	• Archived data related to emerging technology news • Focus on future-oriented technology and breakthrough • Well categorized by technical topic or author, providing detailed information about the author
FutureTimeline	• Articles are sorted by specific forecast timeline • Provides future timelines based on current trends, long-

	term environmental changes, technological development trends, geopolitical evolution, etc • Discussions about timeline creation can be made through scientists, future predictors, and anyone interested in future trends
SingularityWebBlog	• An open community with news and columns on the future of technology and its changes
MIT Technology Review	• Technological analysis magazine published by MIT • Provides an analysis of future-oriented factors of innovative products such as smartwatch and electronic vehicle • The articles are divided according to technical topics, and the technical knowledge level covered in each article is high
World Future Society	• An international community of futurists and future thinkers • People interested in the future freely talk about various futuristic topics
Wired	• Articles focused on how emerging technologies affect culture, the economy, and politics • Offers a wide range of quality journals for a variety of ranges and types of technologies
io9	• focuses on the subjects of science fiction, fantasy, futurism, science, technology and related areas • A type of blog that covers not only professional writing on promising technologies but also related popular culture.
Reddit	• American social news aggregation, web content rating, and discussion website • In the form of 'Subreddit', people with common interests can share their opinions freely

3.5 Conclusion

This study aimed to evaluate the online community to select data sources for public-based technology foresight. To this end, we proposed a framework for evaluating online communities including technology-related content, based on two criteria: expertise and diversity. Specifically, text data relating to the future of emerging technologies was collected from online communities of various types. After that, four indicators were calculated ratio of author per document, optimal number of topics, level of knowledge, and ratio of activists. Based on these, the expertise and diversity scores were finally measured. For both criteria, online communities that meet or exceed a certain threshold were identified as ultimately a valuable data source. In order to verify the measurement method presented in this study, we applied the framework to 20 actual online communities.

The contribution of this study can be summarized as follows. First, the evaluation framework of the online community was newly proposed for the specific purpose of technology foresight. Traditional online community evaluation techniques focused on individual web pages and users rather than the community itself. This is because the purpose of conventional evaluation methods is to derive results from search engines rather than to evaluate the website as a research data source. In response, this study suggested a brand-new evaluation technique for the purpose of assessing data sources for technology foresight. This is expected to contribute to future research as new concept of open foresight emerges.

Second, this study proposed a combined evaluation technique that utilizes a bibliometric approach and text analysis. A bibliometric approach based on the statistics of documents and authors within the online community allows indirect reflection of the size and user characteristics of the website. In addition, text analysis was further conducted to allow consideration of not only the external characteristics of the online community but also the contents contained therein. The combination of this approach is based on both qualitative and quantitative analysis. Therefore, in the long term, it is expected to automate the evaluation of online communities, and through this, new derivative research such as recommendation algorithms of data sources will be possible.

On the other hand, the following limitations exist in this study. First, the number of cases should be increased by applying the framework suggested in this study to more online communities. One of the key differentiators of public-based technology foresight from traditional expert-oriented approaches is the ease of use of big data. In order to specialize these advantages, more data set should be utilized to realize collective intelligence through big data. In this study, we surveyed 20 online communities, but countless online communities can be analyzed. If various websites not considered in this study are explored and analyzed, more valuable data sources can be identified for technical foresight.

The second limitation is that more detailed criteria and indicators are needed for the evaluation of the online community. In addition to the measures used in this study, there are a variety of techniques that can evaluate their effectiveness considering the characteristics of online communities. For example, one method can be to measure how much discussion is done in the form of replies in one thread. In other words, the communication of people is more active in the case of various discussions about a single topic or opinion than when each post exists independently. In addition, connectivity between communities may be considered. Further, for online communities that provide demographic information about users such as gender, age, occupation, and educational background, new evaluation techniques may be developed.

Chapter 4

Public Understanding of Technology

4.1 Introduction

Research on the relationship between public and emerging technologies has been done in many ways. In particular, the public's attitude toward new emerging technology has been emphasized because of its growing influence on their support for the technology, and even the scientific development [56]. Traditionally, in examining public attitudes toward new scientific discoveries researchers have usually relied on two different variables: a measure of general support for or opposition to the new technology and a measure of perception of risks versus benefits of the new technology [68]. These studies have been made under a key assumption that people use their current knowledge about science in order to decide whether they support and adopt it or not.

Recently, however, research on public attitudes toward emerging technology have gradually shifted from a narrow focus on scientific literacy to a more integrated approach in order to understand opinion formation about technology-related issues [98]. The influence of knowledge level as a determinant of public opinion is decreasing since people now form opinions and attitudes even in the absence of relevant scientific or policy-related information [78]. Instead of professional knowledge, they use cognitive shortcuts or heuristics to form judgments about emerging technologies. In line with this, interest on emerging technology might predispose people to seek out information that lead to a higher familiarity, and eventually the public might also be associated with more positive attitudes toward emerging technologies [59, 89]. Namely, the development of new technology is now influenced not

just by the knowledge level, support level, perceived risk or benefit of people but also by the simple interest itself on scientific issues. However, most studies in the public's interest toward emerging technology have only been carried out in a small number of areas. In most of the previous studies on public opinion, there were few cases that quantitatively analyze change of public interest over time.

Yet, Previous researches that focused on the public perception of emerging technologies mainly depended on survey and interviews of experts, or counts of technological documents. Questionnaire and interview data may not be suitable for research to observe the trend of changes over a long period of time and difficult to analyze various technologies due to quantitative limitations. Also, trend analysis through documents counting has a limitation on catching the semantic change of a specific subject.

Meanwhile, due to the rapid development of information and communication technology, large amounts of textual information and knowledge which are written by laypersons became accessible. Many people share their opinions about science and technology or ask others for their ideas via online for various purposes. Besides, one of the main channel of interest in laypeople's technology is internet and online community [89]. They not only acquire information online, but also constantly exchange opinions and communicating about new technology and science [102]. As the consequences of these communications continue to accumulate, it provides a new opportunity for studying the opinions of the general public about emerging technology [7].

In response, this research proposes the framework that explore how public's interest on emerging technologies have been generated and how they change over time by utilizing textual web data. To do this, issues that are related to emerging technologies are extracted by utilizing topic modeling technique. Especially, Latent Dirichlet Allocation (LDA), which is the most popular topic modeling methodology is adopted to time-series database. In other words, time-series topic modeling is utilized to semantically analyze changes in laypeople's main interests for emerging technologies [116]. It is effective in estimating trends of topics extracted from document collection over time and able to figure out the relationships between topics in different time units.

The remaining part of the chapter proceeds as follows. Backgrounds deals with theoretical background of interest of the public and public attitudes on emerging technologies. Next, the proposed approach of this research is described and the result of illustrative case study is also provided. Lastly, the contribution of this study along with future research directions is also discussed.

4.2 Backgrounds

4.2.1 Understanding Public Interest

Interest is a subject that has long been studied in areas such as psychology and education. Although numerous terms are used to describe interest, it can broadly be defined as self-sustaining motives that lead people to engage idiosyncratically with objects and ideas for their own sake [96]. Interest is mainly created by novelty, but meaningfulness and involvement also helps to build interest [83]. After formation, attention is developed by interacting with other groups or individuals [79].

With regard to emerging technology, there are attempts to determine the relationship between interest and attitudes to technology. In the past, knowledge level was important as a determinant of attitudes toward science and technology, but nowadays, the influence of knowledge level as a determinant of attitude to new science and technology is decreasing [94]. Even if they do not have any professional knowledge, people tend to form opinions and attitudes about science and technology. For example, Lee et al. [68] showed that emotional heuristics moderate the effect that knowledge about nanotechnology has on people's overall attitudes toward nanotechnology. Retzbach et al. [89] showed that interest in science and technology is related to positive views on emerging technologies. This is because people who are interested in certain emerging technology are likely find more information about the new technology and become more familiar with it [98].

Methods of measuring the interest include self-reports, neuroscientific methods, and behavioral measures [88]. Self-reports, which is the most common form of interest measurement, allow participants to assess or comment on the degree of interest they experience in a given situation. Neuroscientific methods try to study interest by using human brain activation, but it is still in the early stage. Behavioral measures include observation, ethnography, and online measures. In particular, much of the information available today on the Internet presents new possibilities for analyzing the interests of certain populations [7]. Rech [87] shows Google Trend, web based tool, can be used to analyze the media attention,

search interests, and cause-effect relations of various topics in software engineering.

In this study, we will examine how the interest of user group in the online community changes over time. This is because analyzing how individual interests change over time is both difficult and costly. The way to measure the interest of these groups is to analyze the various documents generated within the online community with the behavioral measures described above. Since the unspecific majority is regarded as the subject of interest and is based on a large amount of data, it is expected to be easier to analyze the process of interest generation and change over time.

4.2.2 Public Attitudes on Emerging Technologies

In technological fields one can often observe waves of public attention combined with high rising expectations on technological possibilities. Such public perceptions, attention, interest, or hype play an important role in the emergence of technology by guiding research activities, attracting resources and creating legitimacy. Such attitudes can be defined as "real time representations of future technological situations and capabilities" [14]. Along with positive promises and hopes of future capabilities, fears and concerns about future risks are parallel features of these kinds of dynamics. Both positive expectations and fears of risk—though different in character and having different dynamics—can be seen to have considerable influence on the discussion of technological change [100]. Now, many of these temporal variabilities in expectations have become part of a widely shared cultural and social stock repertoire for interpreting socio-technical change.

In this regard, Hype Cycle, which is a graphical representation of the ups and downs, peaks and troughs of technological expectations, plays an important role to reflect public attitude to emerging technologies [27]. It is a powerful tool to get the overall grasp of technological innovation development trend, to get the objective assessment of the maturity of technological innovation.

Yet, some limitations exist in the Hype Cycle model. First, the model is too general in not providing enough room for the kinds of variation and unpredictability that characterize the place of expectations in technological social change. The model does not

take into account the diversity and uncertainty of the technology, so it may not reflect the nature of the social change of the emerging technology. In other words, this model assumes that the pattern of change in expectations for all technologies follows same curve. While the existence of hypes is widely recognized in STI-studies, case studies on hypes have thus far remained localized, explaining specific dynamics in specific contexts [104]. Second, Hype Cycle is a qualitative decision-making tool and it relies mainly on the judgment of experts [19]. Thus, a lack of empirical evidence for each stage of the cycle may exist and there more clues for the model is needed. Third, quantitative indicators for the model highly rely on superficial indices of technologies such as the number of participants, the number or the ratio of the technological innovation documents, patent statistical data and the search flow of search engines [27]. There is also a need for a semantic approach that can reflect the actual opinions of users and the public in detail.

In line with this, previous researches that quantitatively analyzed changes in public perception of technology using online data are as follows. He et al. [49] analyzed the change of the scientific topic generated from a large amount of textual data to understand how topics in scientific literature evolve. They suggested iterative topic evolution learning framework by adapting LDA model to the citation network of research paper. However, the data used in this study was limited to research paper and online community data that reflects public opinion does not have an accurate citation relationship. Suggested framework was applied to the ICT industry, but lacking in-depth analysis of semantic contents in the process of change. Zhang et al. [116] developed term-based science map that identify and visualize the evolutionary pathways of scientific topics in a series of time slices. They utilized patents and academic proposals to visualize the time-series snapshots of science map. Maps are term-based but not semantically analyzed at term-level.

4.3 Proposed Approach

As summarized in Figure 4.1, the whole process of the framework presented in this study consists of the following four steps: Construct database, extract technological topics, identify relationships between topics, find pathways of changing interest. The following is a detailed description of each step.

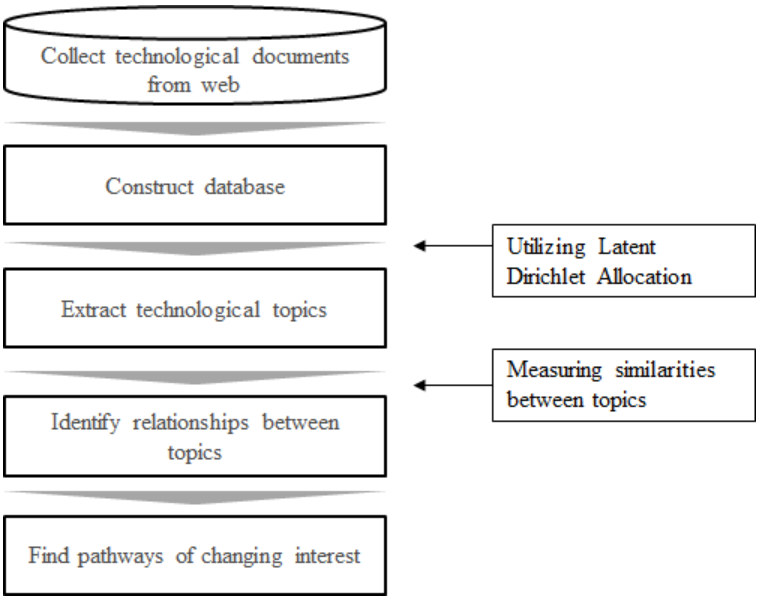

Figure 4.1 Overall framework

First of all, documents from users in the online community with their opinions about emerging technology are collected. The contents contained in the document collection are the raw material of interest to be analyzed in this study. To analyze the change of interest over time, the collected documents are divided by year based on the time they were written. Since the document collected on the web is unstructured data without a certain format, text preprocessing is required. This process includes part-of-speech tagging and stop word processing that remove unnecessary words. To do this, we utilize the Natural Language ToolKits (NLTK), the largest natural language processing project.

Second, we utilize the LDA technique to extract each topic from the previously constructed database. Each topic extracted here corresponds to the interest that people have about emerging technologies. As a result of LDA, topics and the topical keywords that describing each topic are extracted. The keywords are analyzed qualitatively in order to classify and name the topics. Since the database is constructed by year in the previous step, various topics and corresponding keywords are extracted according to the year as shown in the left part of Figure 4.2.

Third, to explore how the contents of each topic change over time, relationships between topics are identified. To do this, similarities between topics in different years are measured. Since the topic extracted through the LDA consists of the probability of topical keywords exceeding a certain threshold, similarity is measured depending on this probability. There are various methods for measuring similarity between topics. In this study, the cosine similarity between topics are used and the relationship between them are identified [57].

Fourth, based on the previously identified similarity between the topics, pathways of the change of the topics are verified. Even topics on the same issue may have different contents over time and these changes can be identified by examining their topical keywords. This makes it possible to understand how the semantic contents of a topic change over time or when the topic itself comes up and disappears as shown in Figure 4.2.

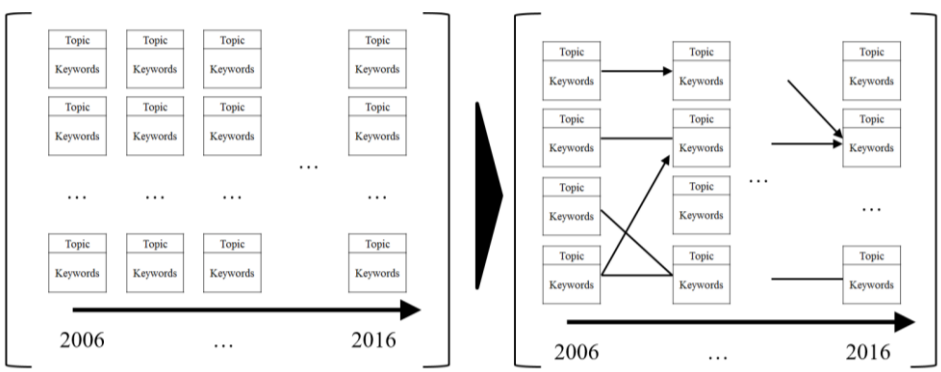

Figure 4.2 An example of finding a pathway of interest

4.4 Illustrative Case Study

This section introduces the results of a case study that applied the framework presented in this study to real data. A series of results from the collection, processing, analysis, and interpretation of real-world data from an online community where public opinions related to various emerging technologies are shared is explained sequentially.

4.4.1 Data Collection

Data for the case study was collected from online community called FutureTimeline. Within the FutureTimeline, users are free to share their opinions on various topics related to science and technology, and within each post, additional communication is actively facilitated through comments. FutureTimeline is relatively public-oriented compared to other technology-related online communities. In other words, the proportion of professional writing such as news and columns is low, and the opinions and predictions related to the technology of non-experts are candidly shared. The structure of the forum is similar to that of Reddit, but the topics covered are much better suited to this study. Figure 4.3 shows an example of discussions and predictions among members on the FutureTimeline.

In this case study, we collected the documents of the category "Science & Technology of the Future", which contains the most topics related to emerging technology in various fields. The articles posted from 2011 to 2017 were collected and a total of 2,258 data were gathered.

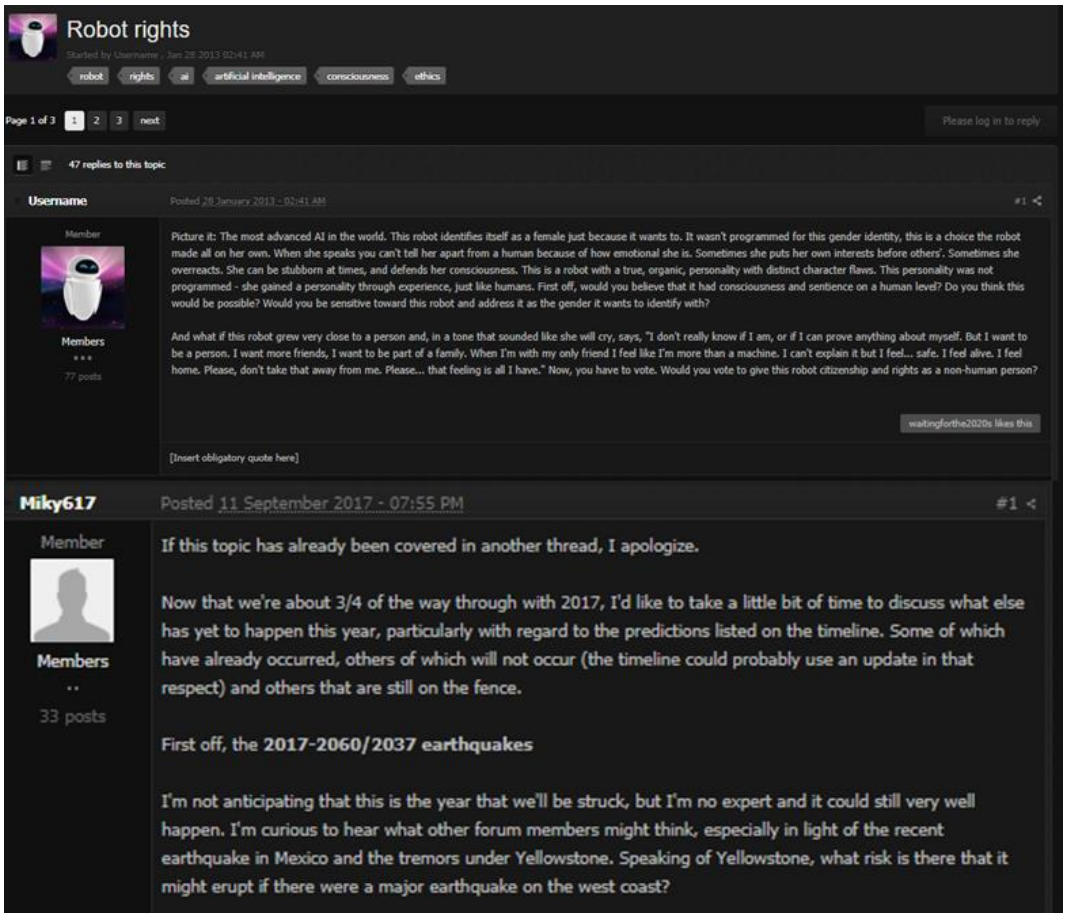

Figure 4.3 Examples of articles of Futuretimeline

4.4.2 Topic Extraction

Next, topic modeling is applied to the data collected for each year. As a preliminary work on this, since the data collected from FutureTimeline is unrefined unstructured data, preprocessing for text analysis was performed. The unnecessary words such as stopwords are removed and the term-document matrix is derived.

After that, topics were derived by applying LDA to the yearly data. To apply LDA, the optimal number of topics must be determined in advance. Several approaches are available to determine the optimal number of topics, and this study refers to four metrics

developed by Griffiths [42], Cao et al. [17], Arun et al. [5], and Deveaud et al. [28]. Each metric has different pros and cons and may furthermore be more or less applicable given the specifics of data. In this study, all of four measures were used to select the most common optimal numbers.

For example, Figure 4.4 shows the result of obtaining the optimal number of topics for data of 2014. The indicators of Griffiths and Deveaud are optimal at higher values, while the indicators of Cao and Arun are optimal at lower values. Therefore, when applying LDA to data of 2014, it is determined that the optimal number of topics considering all four metrics is 7. In this way, the result of determining the optimal number of topics for each year is shown in Table 4.1.

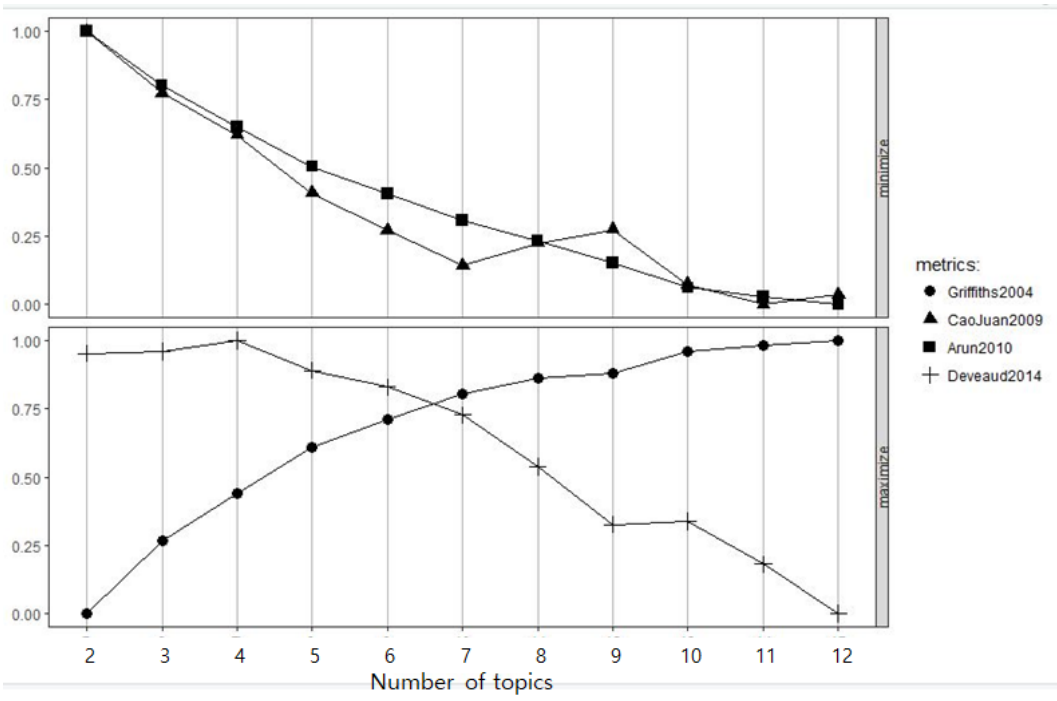

Figure 4.4 An example of finding optimal number of topics

Table 4.1 Summary of topic extraction

Year	Number of docs.	Number of topics extracted
2011	131	8
2012	224	7
2013	514	8
2014	741	7
2015	341	6
2016	336	7
2017	240	8

After applying topic modeling to the data for each year, the keywords constituting each topic and the importance of the keywords are derived. The analyst then examines the keywords of high importance in each topic and names it. For example, if the terms "energy", "cars", "carbon", "electric", and "batteries" are derived as keywords of high importance, the topic can be named "electric vehicle". As a result, the list of topics and keywords with high importance derived for each year is shown in Appendix x.

4.4.3 Identifying Relationship Between Topics

The next step is to measure the similarity between topics from different years. Identifying the links between topics with high similarity to each other, or related topics, is the basis for the final pathway.

In this study, we used cosine similarity to measure relationships between topics of different years [2, 116]. Applying LDA creates a semantic space with topic keywords for each topic. By definition, each topic is a probability distribution over the words in the training corpus. The Euclidean distance between different topics is calculated based on the vectors of the generated topic keywords. The similarity between two topics is calculated as the average pairwise cosine similarity between their top-10 most important keywords.

Figure 4.5 shows the connections between topics from different. For some topics, there were topics that correlated with all years, resulting in a seamless pathway. However, some do not have connectivity with topics of any other years.

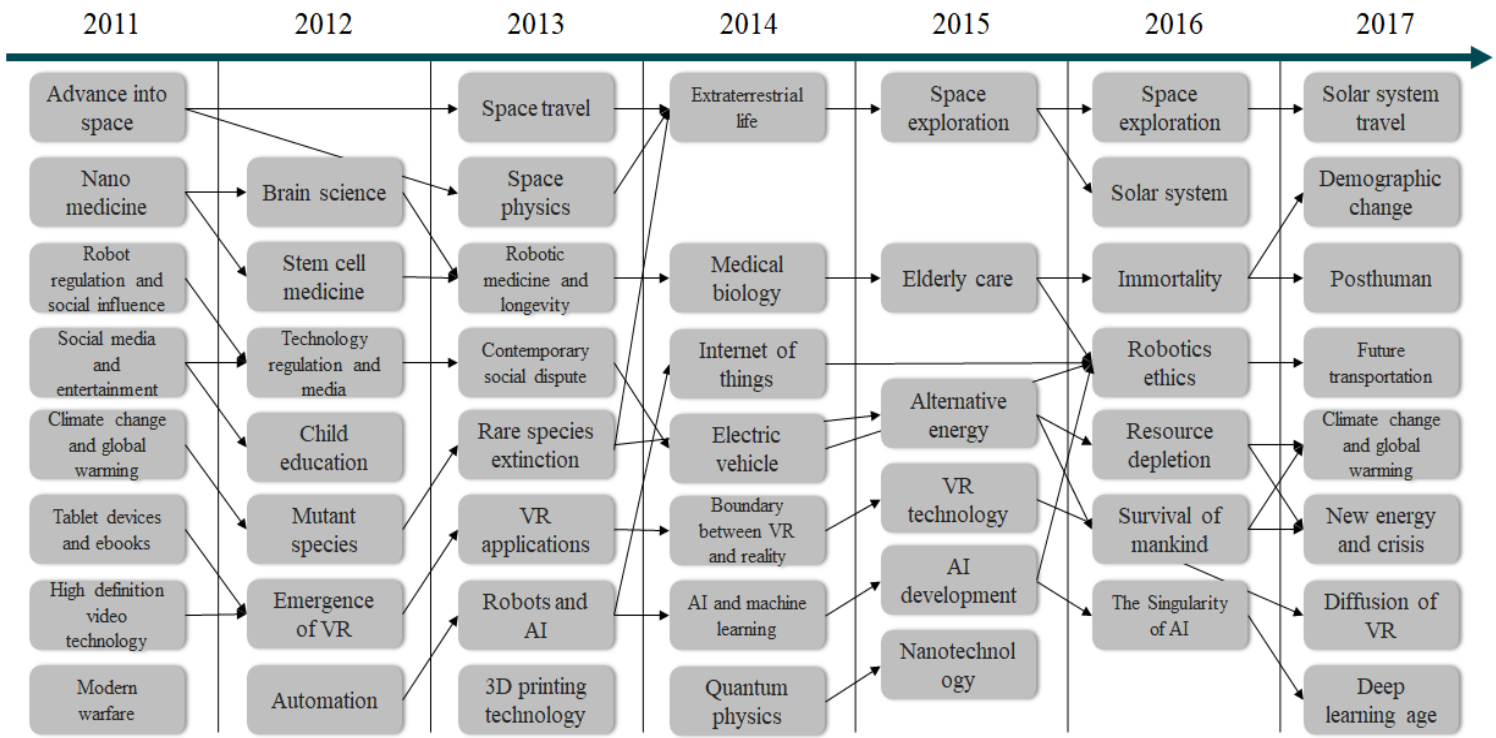

Figure 4.5 Overall linkages of topics

4.4.4 Finding pathways of Changing Interests

Based on similarity between the topics calculated in the previous step, final pathways of topic changes were derived. Namely, the flow of changes in people's perceptions on a specific subject was confirmed through the connection of topics. Unnatural linkages were partially removed at the analyst's discretion. As a result, a total of six pathways were derived and their explanation is as follows.

Figure 4.6 shows the pathway for the VR topic. In this pathway, the process of change of topic is most prominent. It starts with the topic of video technology and portable devices, which are the predecessor of VR, and leads to the emergence of VR. Since then, it has changed into the issue of VR application and related devices, the distinction between reality and virtual reality, and the spread of VR by each year. It can be seen that this flow has a similar flow of technology life cycle model. Especially noteworthy, many skeptical keywords about VR appeared in 2014, which could correspond to the 'trough of disillusionment' phase of Hype Cycle.

Figure 4.7 identifies pathways for topics related to AI. AI-related topics have been steadily increasing since 2012. This pathway was initially started with automation issues. Over time, issues such as singularity and ethics have emerged regarding AI awareness. In 2017, it was developed as an issue related to 'brain'. This suggests that the convergence of AI and brain science could be accomplished later.

Figure 4.8 shows the pathway of topics related to space. Space-related topics were most consistently mentioned among all topics. In other words, we can see that people's interest in the space was solid Notably, generally similar terms are repeated continuously over time. It implies that the interest in the universe is steady, but the level of knowledge itself does not seem to change much. In terms of content, over time, there are more stories about the possibilities of space as a new residence or tourist destination.

Figure 4.9 shows the pathways for topics related to medicine. In the early years, opinions about relatively specific medical fields were given such as nano-medicine, stem cell, and brain science. Recently, the topic has changed to the topic of longevity, furthermore,

immortality and new human. Keywords such as demographic changes in humanity and ethical issues related to these changes are also extracted.

Figure 4.10 is a pathway of topics on social change due to technology and innovation. The process of change is the most complex. There were a lot of opinions on social media and media changes in 2011. It is inferred that this is due to the proliferation of smart devices at that time. By 2013, it has led to the theme of social issues related to technology such as regulation and ethics and the dark side of modern society. In 2014, it led to issues related to IoT and electric vehicle. In 2016 it became a matter of ethics probably due to security and privacy issues.

Figure 4.11 identifies pathways to topics related to the environment. The course of the topic change is various. Pathway's beginning was an issue with climate change, which developed into a topic about mutations and rare species. As time went by, it changed to topics related to the long-term survival of mankind such as alternative energy and resource depletion. Recently, there are more negative keywords such as crisis and survival. In other words, we can see that people's perceptions of the environment have become increasingly pessimistic.

Finally, Figure 4.12 shows topics with no distinct pathway. The reasons why these topics do not show a clear connection with the topics of other years are as follows. First, some topics are newsworthy at specific times. In the case of the modern warfare topic, there were some issues related to the Libyan civil war in 2011. Second, some topics may require a longer term perspective. Topics related to 3D printing or nanotechnology are still important issues in the scientific world at this time, but it may take more time to capture distinct changes.

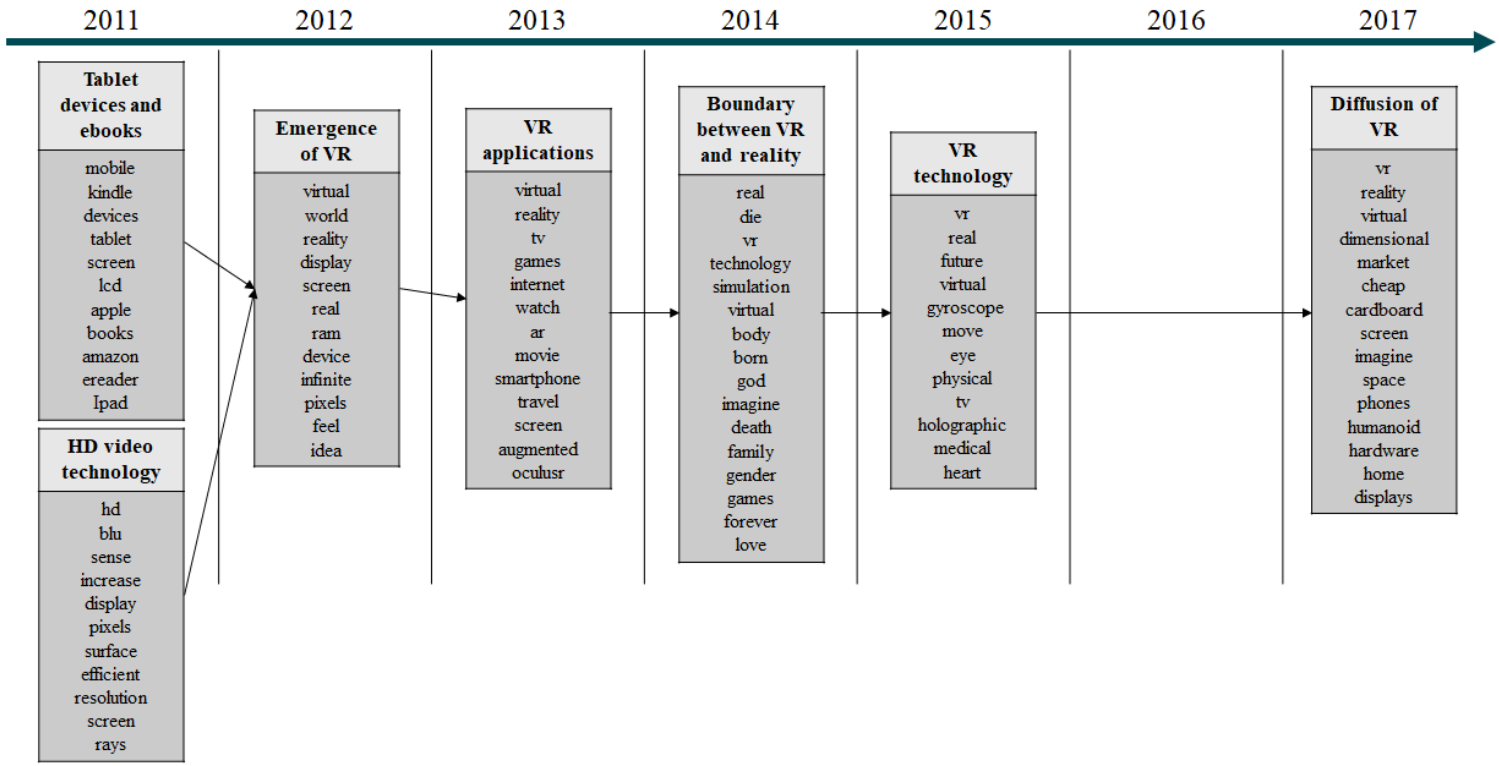

Figure 4.6 VR topic pathway

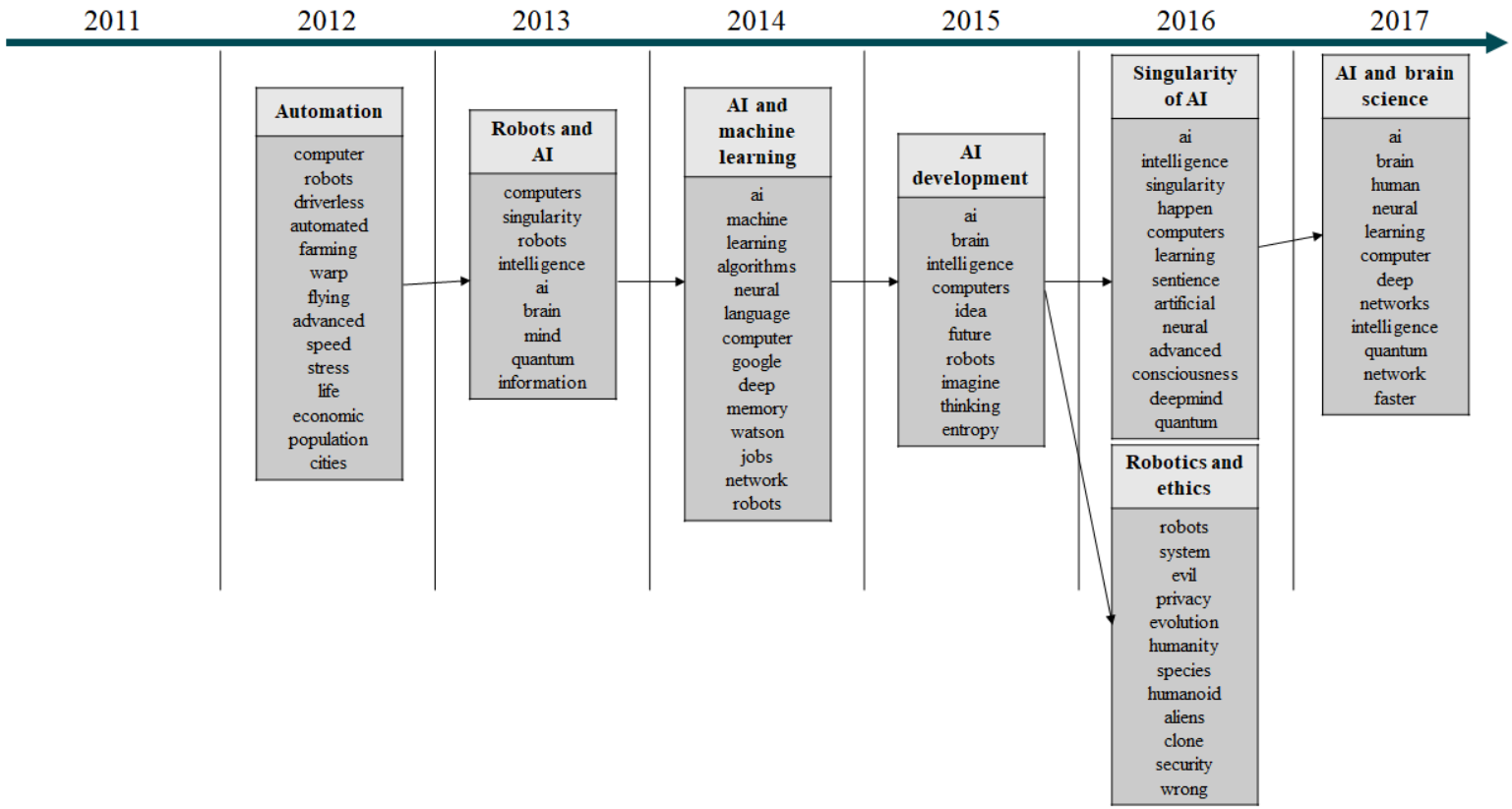

Figure 4.7 AI topic pathway

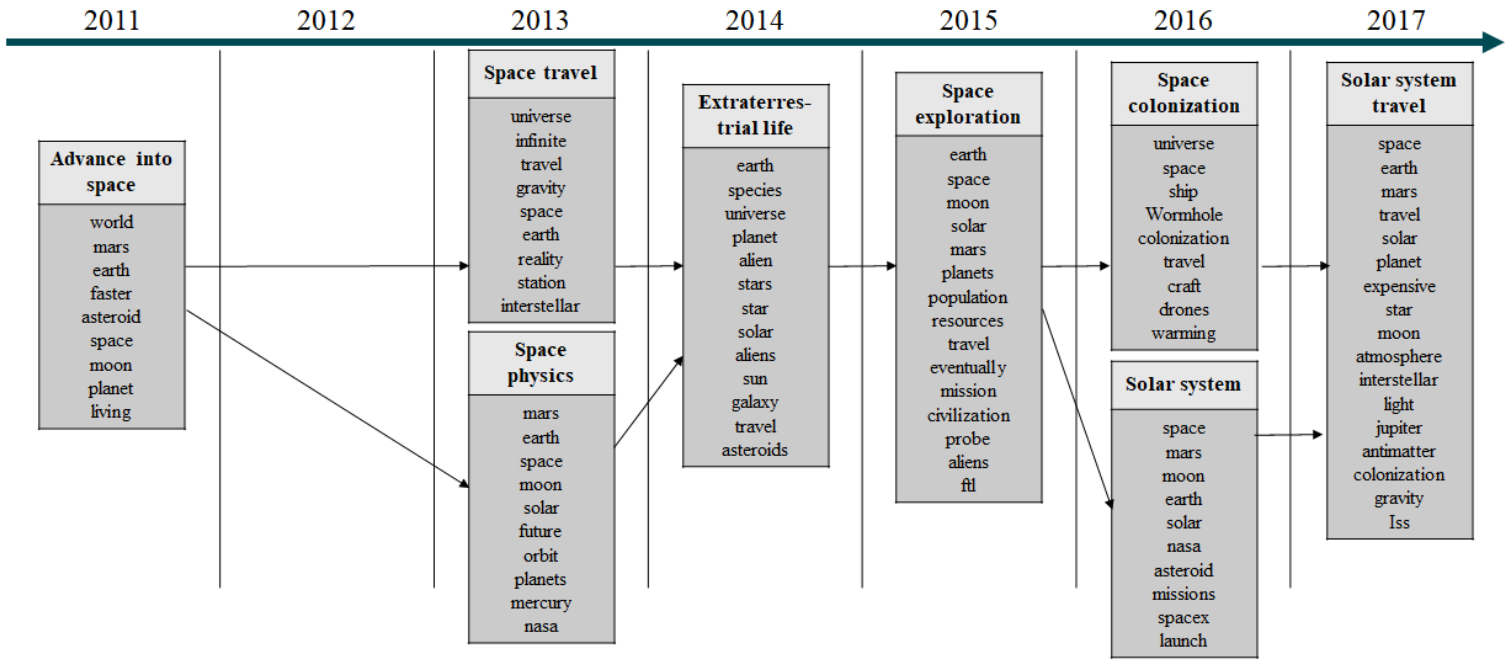

Figure 4.8 Space topic pathway

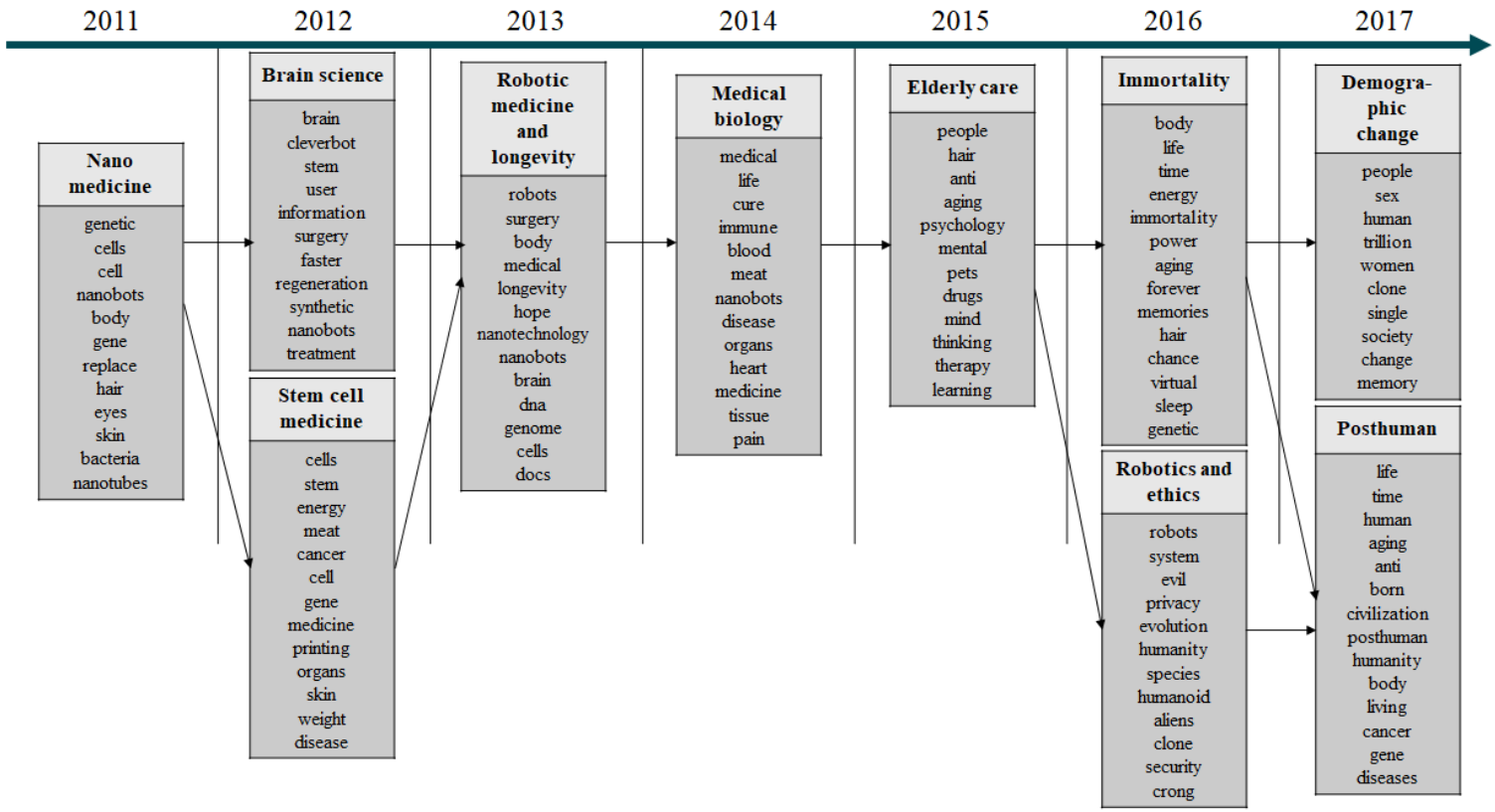

Figure 4.9 Medical topic pathway

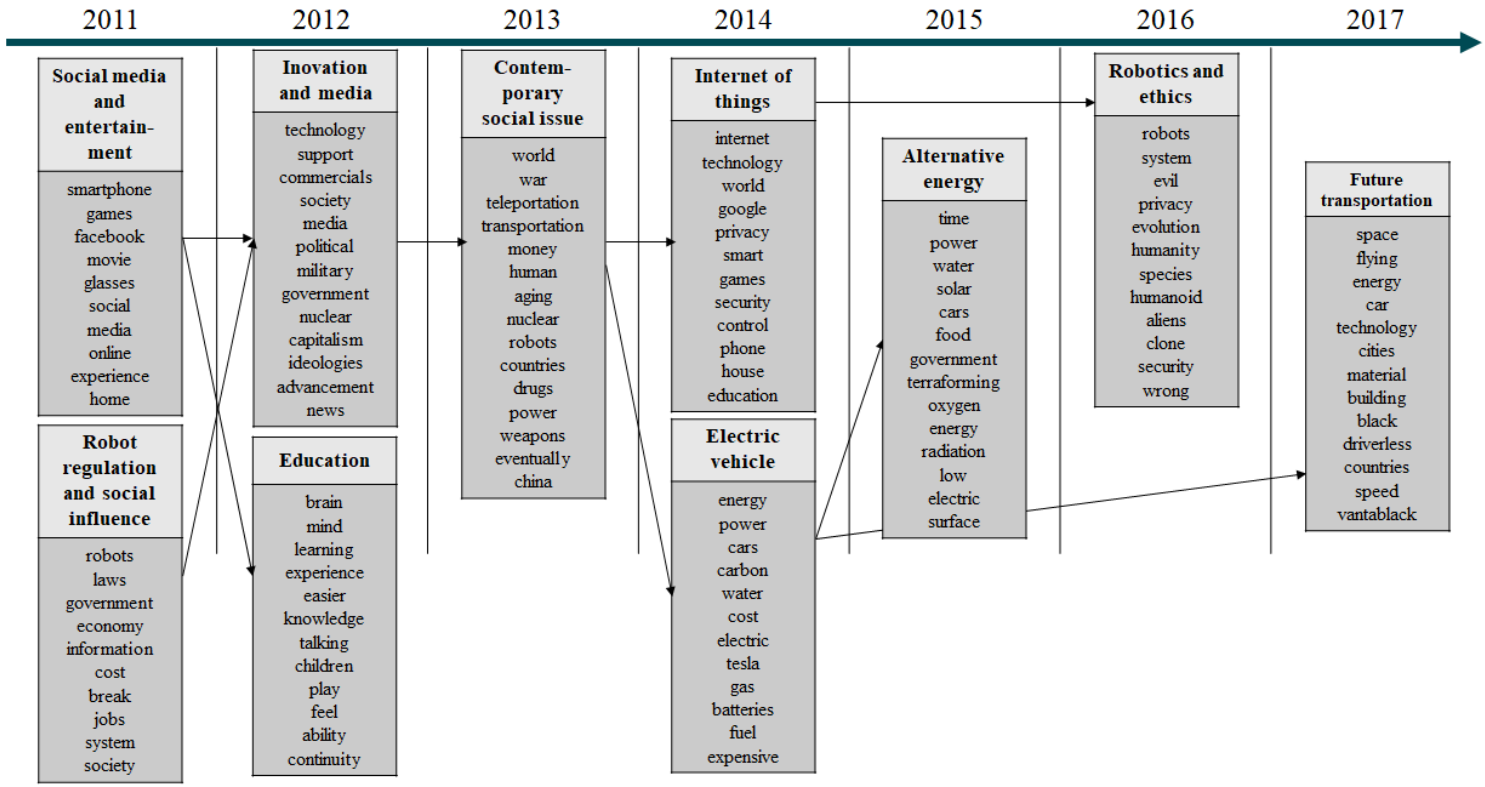

Figure 4.10 Socio-technical topic pathway

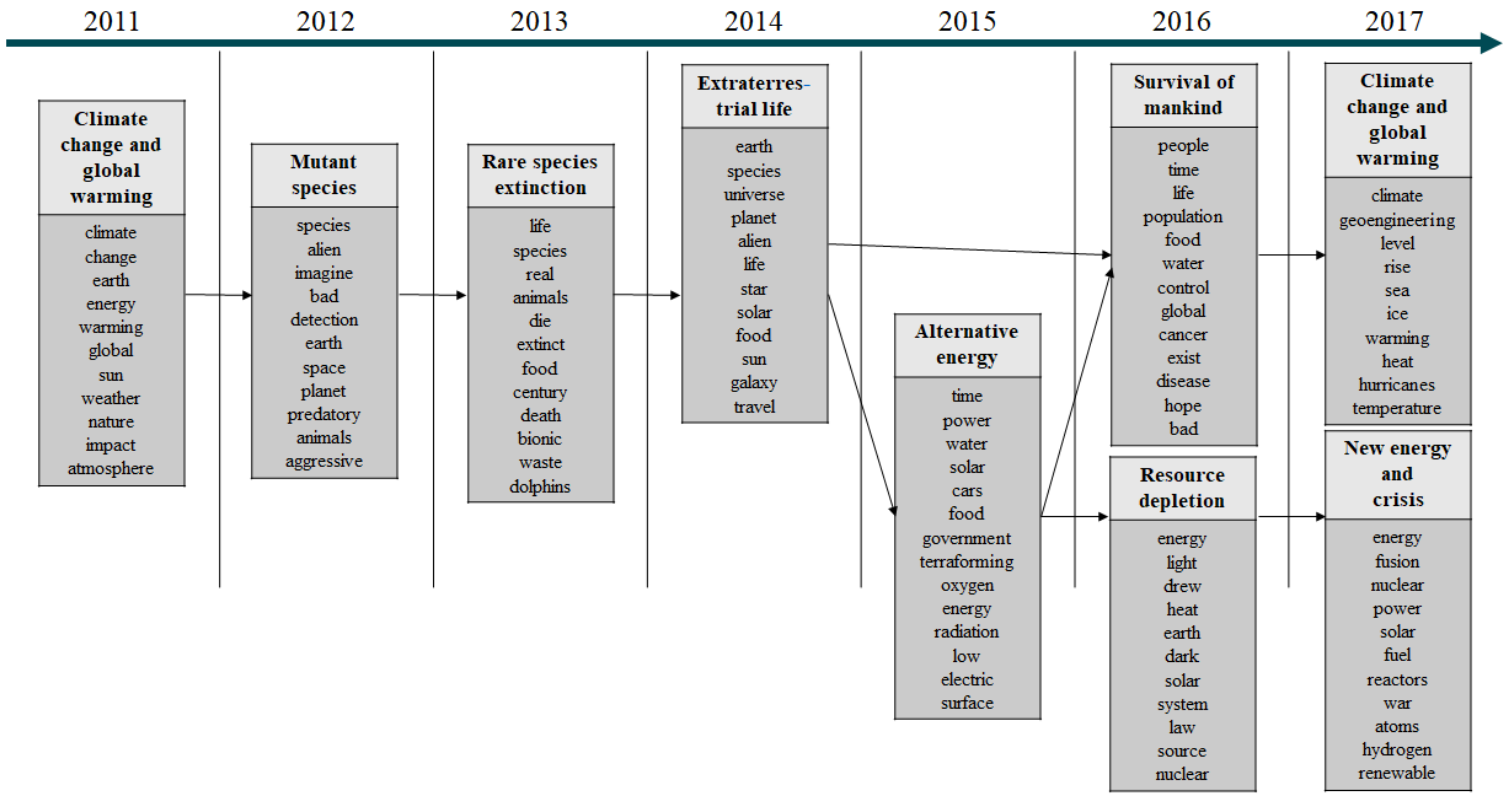

Figure 4.11 Environment topic pathway

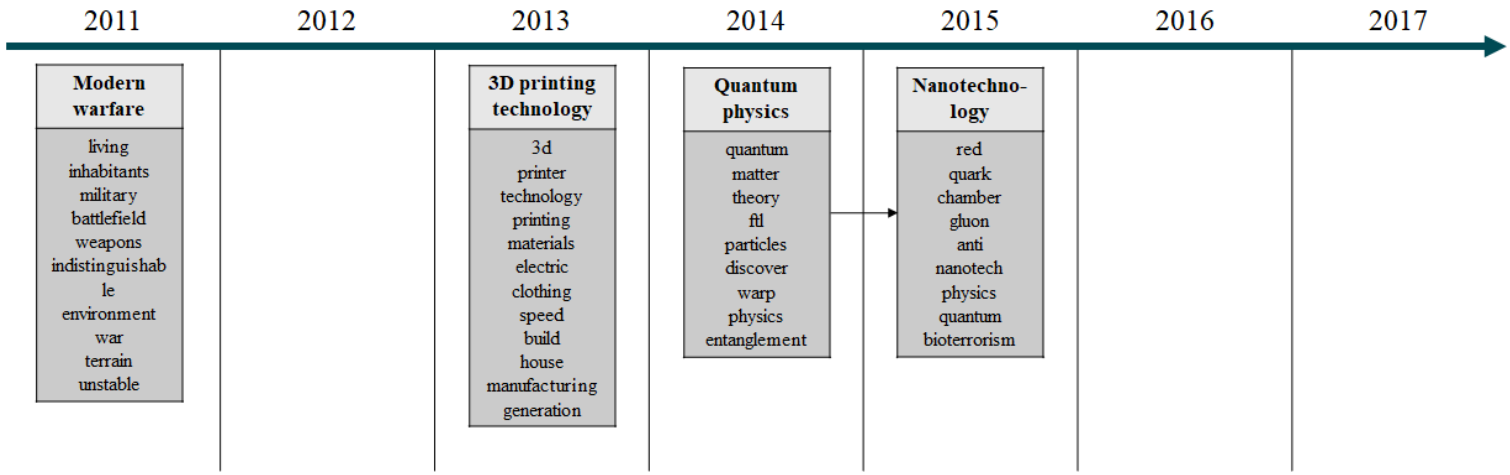

Figure 4.12 Rest topics

4.5 Conclusion

This study proposed a framework that analyze interest of the public on emerging technologies. Textual web-based documents are utilized as a source of people's interest on emerging technologies. In detail, technology-related issues are extracted from online documents by making use of topic modeling technique, LDA. To explore the semantic changes of extracted topics over time, database is constructed by year and relationships between topics are measured based on similarity of their topical keywords. Then pathways that reflect change of public's interest on various technology fields were derived.

The following conclusions can be drawn from this study. First, it was found that the change of topic contents was relatively clear according to each year. Even with topics of the same issue, we could grasp the semantic change of topic contents through topical keywords over time. Second, it was able to see the process of developing on various topics within one pathway. In other words, the analysis of online data shows that there is some similarity to the existing technology-related models such as technology life cycle or Hype cycle.

Contributions of this study can be summarized as follows. First, this research suggests a way to quantitatively measure the public's interest toward emerging technology. Conventional researches that follow the trends of emerging technologies heavily depends on patents, publications or standards [23, 63]. In this study, web-based textual documents, which is written by numerous people, are utilized rather than the scientific literatures or individual survey. Therefore, unlike previous studies which had difficulty in directly reflecting the opinions of the public, it was possible to analyze the perception of general users. Second, text analysis was performed over time through dynamic keyword-based approach. Traditional approaches of topic analysis are relatively static; they apply fixed models to the entire dataset and ignore any possible change resulting from time changes over time [101 ,109]. In response, this study adopted time-series topic modeling technique, which is effective in estimating trends of topics extracted from document collection over time. In this way, it is possible to not only identify publicly focused issues but also explore

the semantic change process simultaneously.

However, there exist some limitations on the research. First, there is a need for case studies with larger scale of database. Since only one community is used as the source in the case study, the diversity of contents is limited. As a result, the number of pathways of perception change finally derived was not that various. Therefore, more diverse and in-depth analysis would be possible if the data of the larger online community is utilized or if the similar communities are integratively considered. Second, it is necessary to refine and verify the method of establishing the linkage between topics. In this study, topics in different time were linked based on one metric, cosine similarity. Adopting other topic similarity metrics may also produce meaningful results. For example, utilizing Kullback-Leibler divergence or Jensen-Shannon divergence may be helpful in better understanding the similarity of more semantic content than cosine similarity.

Chapter 5

Scenario-based Foresight

5.1 Introduction

The technology roadmap (TRM), which enables a firm to carry out R&D activities in a systematic manner in terms of time and method, has long been considered an important tool in technology strategy planning [37]. Recently, uncertainty of future is increasing both in the external environment as well as in the internal environment [53] and it is argued that TRM should reflect these uncertainties and risks. Capturing and predicting changes in the external environment, such as markets, competitors, policies is perceived as a key process for companies to commence technology planning. In addition, the importance of internal uncertainty is progressively emphasized in the technical planning process because of the increasing impact of internal breakthroughs such as technological innovation and product development [64].

Future risks can be caused by such uncertainties, and inadequate responses to them may result in difficulties or even failures for companies [46]. Exploring plausible future through scenario planning is essential for quantifying uncertainty [58, 73] and it has been used for various business strategies including technology planning. Accordingly, one of the most effective approaches to dealing with future uncertainty in technology planning is scenario-based roadmapping, in which scenarios are divided by alternatives of highly uncertain factors and the planning is carried out accordingly. Exploratory scenarios can describe the desired future at an early stage, and the goal-oriented roadmap can be planned as policies and activities over time, so they are complementary.

However, most existing scenario-based TRM studies have had limitations in effectively handling both internal and external uncertainties [58]. First, previous researches focused mainly on analyzing external uncertainties and lacked consideration of internal uncertainties. In other words, internal uncertainties that may arise when selecting and developing products or technologies, as well as changes in external environments such as markets and industries, should be explored. Second, existing researches were short of flexibility in the scenario building process. In previous studies, scenarios were typically set and classified simply as positive, neutral, and negative, or factors and variables that constitute scenarios were often given beforehand. Third, existing related studies have limitations in that the various linkages between scenarios and their strategic options are not considered.

Meanwhile, due to the development of information and communication technology, it is becoming possible to exchange information on future technologies in various fields such as experts, the public, and corporations in real time, which is causing phenomena such as Web 2.0 and collective intelligence [44, 93]. As a result, a variety of big data are emerging, providing new opportunities for technology planning and strategy. Big data-based TRM is able to consider far more perspectives and predictions than traditional TRMs, which are created through the consideration of a few experts. Since big data is generated by the masses and experts across many disciplines, it is possible to explore uncertainties of the future that may arise from the interaction of various business factors. In other words, it has become possible to utilize various big data for TRM, which can facilitate the analysis of various internal and external environments.

In line with this, this study proposes a technology planning framework to cope with future uncertainty by deriving future technological scenarios from big data and constructing scenario-based TRM. In this process, Fuzzy Cognitive Maps (FCM) technique, which analyses the causal relationship between various factors, is adopted and utilized for the quantitative scenario development. While FCM has strengths in scenario analysis such as measurement of scenario occurrence and influence probability or construction, simulation and evaluation of scenarios, TRM is useful for strategic planning to grasp technological

factors inside and outside the organization and support decision making. Therefore, the suggested framework integrates two methodologies into the whole technology planning process. In addition, the big data used in this study is an unstructured data for which a specific format is not defined. Therefore, various text mining techniques are used to analyze the data in order to utilize it as an input of the FCM.

The remainder of this chapter is organized as follows. Literature reviews related to scenario-based technology roadmap are explained. Next, the research framework proposed in this study is described. A case study on UAV and its results is also provided to confirm the usefulness of the proposed framework. Finally, conclusions, limitations and future research of the study are explained.

5.2 Scenario-based TRM

Scenario-based TRMs have been effectively used to analyze future uncertainties in the technology planning phase. The overall objective is to classify scenarios depending on alternatives to high uncertainty factors and plan the future accordingly. Scenario-based TRMs can be effective for the reason that scenarios and roadmaps are complementary to each other [91]. Since the scenario is participative and interactive, it is used to describe various situations. However, the process of analyzing scenarios is not intuitive and can be difficult to understand. On the other hand, the roadmap focuses linearly only on the desired future. However, since the various information is expressed through an intuitive diagram, it can be understood accurately and concisely. Thus, scenarios can be used in the early stages of strategy exploration, and roadmaps can help to achieve a goal-oriented plan by converging the results.

Previous studies suggesting a scenario-based technology roadmap can be classified into two types: a proposal of a new conceptual structure and a proposal of a methodological extension. Research on the conceptual structure has newly proposed roadmap structure, elements, and roadmapping process [30, 36, 91, 99]. For example, Drew [30] proposed a roadmap consisting of long-term macroeconomic factors and trends, such as the geopolitical and socioeconomic environment of future scenarios. Strauss and Radnor [99] defined key activities, interdependencies, decision points that constitute a multi-path roadmap according to scenarios. Contrarily, studies that attempted methodological extension have consolidated various quantitative approaches such as Bayesian networks for evaluating the ripple effect of future change [67], relationship between external environment and internal strategy [38], mutual impact analysis [82].

More recently, research on TRM has expanded in many respects, including subjects, data, and applications. Zhang et al. [115] presented a hybrid TRM that combines qualitative expert knowledge and quantitative text analysis derived from Science, Technology & Innovation data to capture competitive technical intelligence and applied proposed model to dye sensitized solar cells field. Lu and You [71] suggested a roadmap

construction and evaluation framework for defense technology through a multi-layered knowledge flow network model derived from patent and academic data. Cheng et al. [20] proposed a scenario-based roadmap framework that considers both country and industry level macro-views and organizational-level micro-views. In particular, Siebelink et al. [95] proposed a robust roadmap technique to reduce future uncertainty through scenario-driven roadmapping. This was applied to a Dutch construction firm and proved its utility. The main difference from our study is that this research considers environmental uncertainty mainly and that the scenario is based on an expert workshop rather than big data.

In this study, we utilize FCM to construct scenario. The FCM-based scenario planning has the following advantages compared with the existing scenario-based roadmapping. First, FCM can model various factors and relationships by integrating knowledge gained from experts and data. Second, the uncertainty of various factors, both internal and external, can be considered because it includes the concept occurrence probability value. Third, the relationship between various scenario elements and strategic alternatives can be grasped by expressing the positive or negative causal relationship between concepts. Finally, it supports flexible scenario development and simulation.

5.3 Proposed Approach

The overall framework proposed in this study is shown in Figure 5.1.

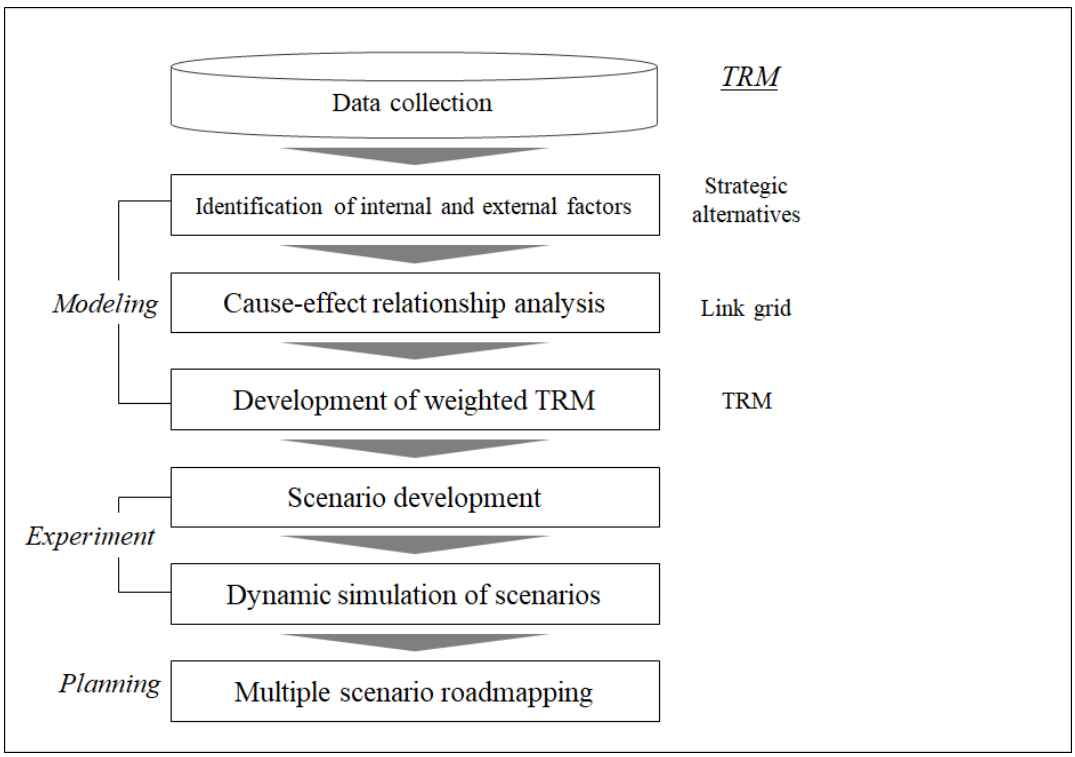

Figure 5.1 Overall framework

5.3.1 Identification of Internal and External Factors

First of all, a database is constructed by gathering various web-based data related to emerging science and technology. To do this, textual data related to emerging technologies is identified and selected and a crawler that can be applied to each data source is developed to automatically collect, extract, and store data. Database is constructed by structuring unformatted data through preprocessing and refinement.

In data preprocessing, topic modeling is performed to identify concepts related to

future technologies in terms of topic. Especially, Latent Semantic Algorithm (LSA), one of the representative algorithms of topic modeling, is applied [31]. In this study, semantic textual patterns resulting from LSA are considered as keyword clusters, and each keyword cluster is used as a concept node of FCM.

These concepts which are internal and external factors in TRM, are used to develop scenarios through FCM. Identified concepts are classified into the TRM layer [53]. Concepts are divided into external factor ‘market’ which is related to business, market, and policy, and internal factors ‘product’ which is related to product, service and capability, and ‘technology’ related to technology, competitiveness and resource. In addition, an 'outcome' layer that can express performance through technical planning and innovation, such as the number of users, profit, and utility is set up. The uncertainty of the concepts is evaluated and the probabilities of occurrence are calculated between 0 and 1. The probability of occurrence of a concept is calculated based on the probability that the corresponding topic is mentioned in the document, and the probability of an outcome is determined qualitatively. For example, the outcomes generated by the development of industries related to the technology can be classified into societal and economical values, and the probabilities of these factors are determined by the analyst's judgment.

5.3.2 Cause-effect Relationship Analysis

At this step, ARM is performed to calculate the causality weights that represent the causal relationship strength and sign between concepts. The three indices of support, confidence, and lift used in association rule mining determine the significance concept, the absolute value of the causal weight, and the sign of the causal weight, respectively [58]. As a result, all concepts corresponding to market, product, technology, and outcome constitute row and column of a causal matrix as shown in Figure 5.2, and the value of each cell has a causal weight between -1 and +1. The blue part of the matrix represents the intra-layer relationship, and the rest represents the inter-layer relationship in the TRM.

	M1	M2	M3	P1	P2	P3	P4	T1	T2	T3	O1	O2
M1		-0.7	0.2	0.8								
M2			0.5									
M3												
P1												
P2												
P3												
P4												
T1												
T2												
T3												
O1												
O2												

Figure 5.2 An example of causality matrix

5.3.3 Development of Weighted TRM

Based on the previously determined concept, the probability of each occurrence, the classification of TRM layers, and causal weights among each other, FCM is drawn on the TRM, resulting in a weighted TRM as shown in Figure 5.3.

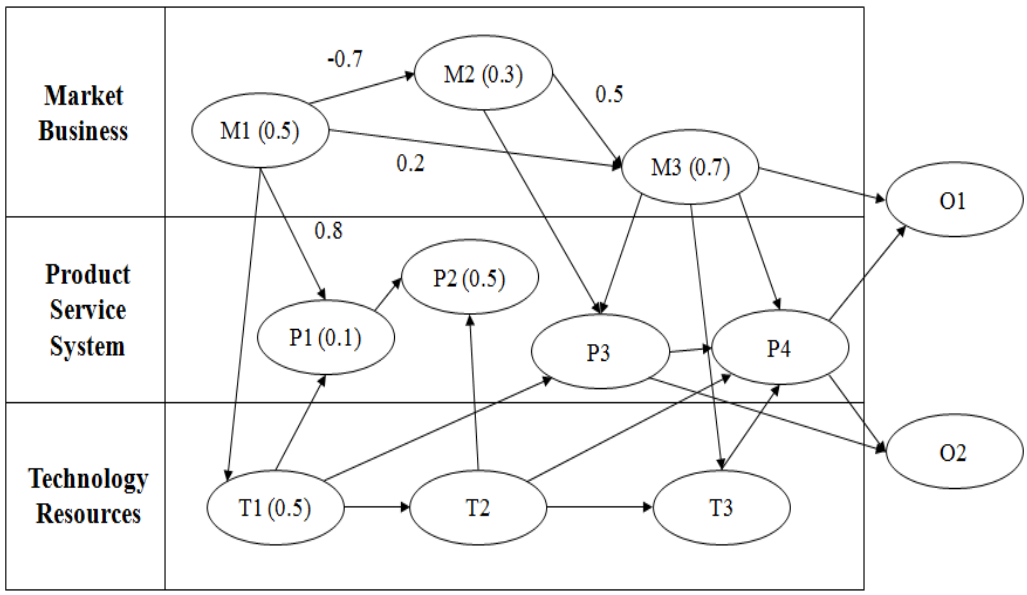

Figure 5.3 An example of a weighted TRM

That is, in the weighted TRM, concepts are expressed in layers according to the form of TRM, and weights are expressed in concept nodes and causal weights are shown on connection lines as in FCM. 'Strategic decision making' is not yet considered in the weighted TRM, which is a preliminary form of the final scenario roadmap.

5.3.4 Scenario Development

In this study, any combination of binary state values of weighted TRM concepts is considered as input vector, which is background or start condition of one scenario. To this end, a morphological analysis, which can be utilized for scenario analysis based on FCM, is adopted [3]. As shown in Figure 5.4, variation can be defined according to the occurrence of each concept, and the input vector is constructed by combining concept changes of market, product, and technology layers.

	Scenario projection			Strategic options			
	M1	M2	M3	P1	P3	T1	T2
Variation A	Increase	Leisure buying	…	Develop	…	Develop	…
Variation B	Stagnation or Decrease	Functional buying	…	Withdraw	…	Withdraw	…

Figure 5.4 Scenario development based on combination of concepts

However, combinations that contain inconsistent pairs among total generated combinations are eliminated, and input vectors are extracted by selecting several strategically most significant combinations.

5.3.5 Dynamic Simulation of Scenarios

The output vector of a given input vector can be drawn via dynamic simulation of FCM. The dynamic simulation is conducted by updating the state value A_i of the concept node C_i according to the time t as follows:

$$A_i(t+1) = S\left(A_i(t) + \sum_{\substack{j \neq i, \\ j=i}}^{N} A_i(t) \cdot W_{ji}\right) \tag{5.1}$$

The state value at $\mathrm{t}+1$ is obtained by applying the limit function to the product of the state value $A_i(t)$ and the causal weight values W_{ji} and converting to a binary value. This process is repeated and stabilises the value of A_i and converges to a certain value. This value is considered to be the final output vector of the input vector.

5.3.6 Multiple Scenario Roadmapping

Finally, decisions based on strategic alternatives are made and roadmapping regarding selected multiple scenarios are implemented as shown in Figure 5.5.

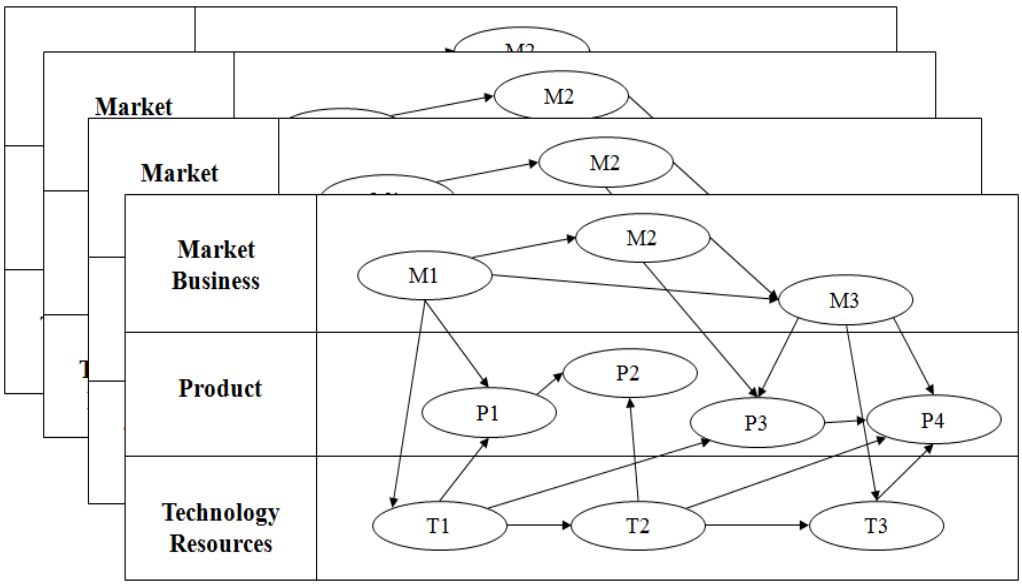

Figure 5.5 An example of multiple scenario roadmaps

5.4 Illustrative Case Study: Unmanned Aerial Vehicle (UAV) roadmap

UAV is a vehicle that is designed to perform a specified mission without boarding a pilot, often called a drone. In recent years, UAV technologies have been developed in many countries around the world. In particular, high-performance UAVs have been researched with the United States and Israel as leading. Countries around the world are developing various technologies related to a drone for military purposes, such as stealth, arms, strategic tactical surveillance, vertical take-off/landing, and supersonic, in the field of unmanned systems. It is the fastest growing field in the aerospace industry and expected to be more promising in the future.

UAVs, which have been developed for military purposes, are now being applied to the private sector. It is mainly utilized for surveillance, R&D, filming, crime investigation, logistics, and communication. Civilian drones are able to collect information by observing the ground from the air and approaching where people are not accessible to collect scientific data. They are also actively used in the field of photography and criminal investigation to

help arrest criminals. In addition, in the logistics business, small couriers can be delivered quickly using small drones, and they are also used for communication signaling.

Although UAVs are useful in many fields, there is high uncertainty in the development of UAV technology. Application of unmanned aerial vehicle technology can lead to unexpected threats and opportunities such as various legal regulations and market changes in future and these factors need to be considered in the process of technology planning. Since UAV technology is one of the representative emerging technologies, the future prediction and discussions of the technology are variously performed in the web environment. Therefore, UAV technology is selected as a case study of the big data-based scenario road mapping of this study.

5.4.1 Data Collection

As shown in Table 5.1, the following five websites, which contain many references and discussions on UAV technology, were selected as data collection sources. A total of 3,236 textual documents were collected from the search results of keywords "drone" or "uav" and "issue" or "future" or "development". Collected documents cover a variety of issues, including social, technical, economic, environmental, and political issues of UAV technology, and consist of unstructured text corpora.

Table 5.1 Data sources for UAV technology

Website	# of collected documents
Future Timeline (http://www.futuretimeline.net)	183
MIT Technology Review (http://www.technologyreview.com)	994
World Future Society (http://www.wfs.org)	678
Wired (http://wired.com)	965

io9 (http://io9.com)	416
Total	**3,236**

5.4.2 Identifying Internal and External Factors

Text mining analysis was performed to extract concepts from collected data. To this end, NLTK (Natural Language Toolkits), which is a representative lexical analyzer, was utilized. First, tokenizing and part-of-speech tagging were performed to determine the parts of all the words included in the data, and 16,086 words were extracted. By applying stop-word processing, meaningless words were removed and only general nouns, verbs, and adjectives were selected. As a result, a total of 2,210 words were extracted and a term-document matrix of 2,210 rows and 3,236 columns was generated based on the frequency of the extracted words as shown in Figure 5.6.

	1	2	3	4	5	6	7	8	9	10	11	12	14	16	17	18	19	20
privacy	1	6	35	35	35	0	0	0	2	2	2	0	2	3	2	2	3	3
have	0	5	20	20	20	5	0	130	6	6	0	6	5	3	16	0	7	7
news	1	13	3	3	3	1	0	26	0	0	0	0	24	5	5	8	0	0
people	1	0	4	4	4	2	0	30	2	2	0	2	1	2	2	0	1	1
first	1	2	5	4	4	2	0	4	1	1	1	9	0	2	5	1	0	0
email	0	13	4	4	4	0	1	0	0	0	0	1	0	6	0	4	4	4
like	1	1	3	3	3	1	0	2	0	0	0	0	1	3	5	6	0	0
see	1	1	0	0	0	1	0	3	0	0	0	4	0	1	1	3	1	1
do	0	1	6	6	6	1	0	11	0	0	1	0	1	0	6	0	5	5
world	0	15	1	1	1	0	0	28	2	1	4	6	2	7	13	3	11	12
share	0	74	2	2	2	1	0	0	0	0	4	0	16	0	1	4	1	1
take	0	1	1	1	1	0	0	5	0	0	0	0	1	2	4	1	0	0
help	0	1	0	0	0	0	0	3	1	1	1	3	1	1	3	1	0	0
follow	0	2	2	2	2	0	0	0	0	0	0	0	1	2	1	21	2	2
while	0	0	8	8	8	4	0	6	0	0	1	2	0	0	1	0	1	1
flying	0	0	3	3	3	0	0	0	0	0	0	0	0	0	1	0	2	2
small	0	2	3	3	3	0	0	2	0	0	0	0	3	3	2	0	1	1
google	1	12	5	5	5	0	0	2	1	1	2	20	2	3	1	0	0	0
state	0	0	3	3	4	0	0	8	0	0	0	4	1	1	7	0	0	0
subscribe	1	2	7	7	7	0	0	0	1	1	1	0	0	0	3	6	0	0
last	0	0	1	1	1	5	0	3	0	0	0	0	2	1	4	1	3	3
report	1	4	2	2	2	4	1	12	1	1	1	2	0	0	4	0	0	0
police	0	18	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
know	1	2	2	2	2	0	1	1	0	0	0	0	1	0	3	2	0	0
issue	0	0	9	9	9	0	0	3	0	0	0	0	0	0	2	0	0	0
comment	0	1	2	2	2	0	0	0	0	0	0	0	0	1	0	2	2	2
content	0	2	0	0	0	0	1	13	0	0	0	1	1	0	1	0	0	0
life	0	39	0	0	0	0	0	7	1	1	1	2	0	1	0	1	0	0

Figure 5.6 Part of the generated term-document matrix

Next, LSA was applied to find semantic textual patterns of the term-document

matrix. As a result, 15 semantic text patterns were derived from the keyword cluster, and the topic was defined as Table 5.2, taking into consideration the meaning of each keyword. These are used as the concepts of FCM.

Table 5.2 Results of concepts extracted by topic modeling

Concepts	Related keywords
Warfare and weapons	agent, capture, border, captain, gear, veteran, cruise, war, better, kill, state, power, oppose, campaign, military, repression, president, force, fuel, imprisonment, murder, secret, squadron, artillery
Political application	battery, quadcopter, surge, tract, record, gear, open, material, engine, fuel, light, element, matter, structure, design, automation, efficiency,
Privacy regulation	statute, legislature, jurisdiction, federalism, regulate, trespass, criminal, conduct, violate, privacy, creepy, constitution, property, dividend, cybersecurity
Terrorism	extremism, terrorist, racist, predator, dead, strike, militant, bomb, enemy, tower, missile, trauma, briefing, guardian, revenge, bomb, propaganda, counterterrorism
Daily surveillance	whistleblower, encryption, population, bounty, rating, offline, truth, screengrab, spying, podcast, investigator, scheme, keyboard, interview
Disaster and safety	police, storm, evacuation, die, coast, surge, travel, hurricane, flood, hazard, cold, instability, violate, vulnerability, border, warrant, prevention, hazard, criminal
Ecosystem disturbance	hummingbird, ecology, detect, carbon, intrusion, sustainability, bug, eagle, species, wilderness, biologist
Logistics	amazon, transport, packaging, docking, delivery, workplace, shipping, shopping, precision, neighborhood, membership, door, subsidy, unemployment
Agricultural support	farming, zoning, mark, crop, imagery, deforestation, forest, screen, verification, pollution, chemical, livestock, tracking, seed, conservation
Entertainment	heritage, programme, entertainment, recording, highlight, gathering, library, briefing, soccer, tourism, storage, auto, voyage, racing, motion, climbing, mini
Detection avoidance	obstacle, control, landing, sonar, carrier, collision, avoidance, robotics, travel, platform, shadow, cruise, awareness, scan, image, drop, altitude, velocity
Internet service	broadband, mobile, sun, warming, cloud, server, glider, wireless, temperature, repeat, planning, guide, density, establish, database

Software technology	hunting, observation, biometric, monitoring, precedent, treatment, improve, risk, development, assessment, analytic, framework, storage, warning, respone, prevention
Navigation technology	pilot, radar, processing, monitoring, access, miss, return, reliability, driving, navigation, air, telecommunication, traffic, gps, control, traffic
Platform and power technology	imprisonment, censorship, injustice, tyranny, repression, plutocracy, genocide, cruelty, prosecutor, campaigning, coup, impunity, punishment, reconnaissance

The extracted concepts were classified into the layers of TRM, market, product, and technology (see Table 5.3). Occurrence probabilities of these concepts were calculated based on the probability that the topic was mentioned in the document. The probability of occurrence is calculated based on the importance of the topic in the document, which is extracted by applying LSA. In addition, the outcome layer was set to economic value and societal value, which are related to the concept of the preceding layer, and their occurrence probabilities were considered to be 1 because the economic and social impact of UAV development is almost certain.

Table 5.3 Classification of concepts and their probabilities

Layer	Concepts	Occurrence probability
Market	Daily surveillance	0.569
	Terrorism	0.798
	Privacy regulation	0.971
	Ecosystem disturbance	0.394
	Political application	0.425
Product	Logistics	0.805
	Agricultural support	0.637
	Disaster and safety	0.551
	Warfare and weapons	0.953
	Entertainment	0.717

	Internet service	0.241
Technology	Platform and power technology	0.725
	Detection avoidance	0.452
	Navigation technology	0.299
	Software technology	0.831
Outcome	Economical value	1
	Societal value	1

5.4.3 Cause-effect Relationship Analysis

ARM was performed to calculate causality weights between concepts. The matrix V, which is the result of the application of LSA, indicates the influence of the topic shown in each document and was utilized as input data of ARM. Therefore, ARM was performed under the condition that the minimum support is 0.15 or more and the minimum confidence is 0.3 or more. As a result, 38 association rules were derived as shown in Table 5.4.

Table 5.4 Result of association rules

Association rules			Support	Confidence	Lift
Navigation technology	→	Logistics	0.33	0.651	1.3
Platform and power technology	→	Entertainment	0.32	0.649	1.29
Entertainment	→	Logistics	0.32	0.647	1.29
Agricultural support	→	Logistics	0,32	0.644	1.29
Software technology	→	Platform and power technology	0.32	0.642	1.28
Platform and power technology	→	Agricultural support	0.31	0.639	1.28
Navigation technology	→	Entertainment	0.31	0.637	1.27
Agricultural support	→	Disaster and safety	0.31	0.635	1.27
Platform and power technology	→	Logistics	0.31	0.633	1.27
Software technology	→	Agricultural support	0.31	0.631	1.26
Disaster and safety	→	Platform and power technology	0.3	0.624	1.25
Platform and power technology	→	Ecosystem disturbance	0.3	0.617	1.23
Detection avoidance	→	Software technology	0.3	0.614	1.23
Warfare and weapons	→	Terrorism	0.3	0.607	1.21

Navigation technology	→	Detection avoidance	0.3	0.605	1.21
Detection avoidance	→	Ecosystem disturbance	0.3	0.603	1.21
Software technology	→	Political application	0.3	0.601	1.2
Logistics	→	Ecosystem disturbance	0.29	0.591	1.19
Navigation technology	→	Internet service	0.29	0.587	1.18
Internet service	→	Entertainment	0.29	0.575	1.17
Platform and power technology	→	Daily surveillance	0.28	0.568	1.17
Political application	→	Daily surveillance	0.28	0.561	1.16
Software technology	→	Daily surveillance	0,28	0.554	1.15
Entertainment	→	Daily surveillance	0.27	0.554	1.14
Logistics	→	Daily surveillance	0.27	0.543	1.13
Political application	→	Privacy regulation	0.26	0.529	1.12
Navigation technology	→	Warfare and weapons	0.26	0.508	1.12
Software technology	→	Warfare and weapons	0.25	0.501	1.11
Terrorism	→	Navigation technology	0.24	0.476	0.96
Privacy regulation	→	Terrorism	0.23	0.45	0.9
Privacy regulation	→	Daily surveillance	0.23	0.441	0.88
Privacy regulation	→	Internet service	0.23	0.434	0.87
Warfare and weapons	→	Entertainment	0.22	0.425	0.85
Terrorism	→	Software technology	0.21	0.418	0.83
Terrorism	→	Agricultural support	0.2	0.407	0.81
Terrorism	→	Entertainment	0.18	0.386	0.77
Terrorism	→	Platform and power technology	0.16	0.376	0.75
Terrorism	→	Disaster and safety	0.16	0.376	0.75

In association rules X→Y, confidence can be interpreted as the influence or strength of the causal relationship because it represents a conditional probability that Y will occur when X occurs. On the other hand, the lift indicates the ratio of cases where Y is randomly generated and X is taken into consideration and it represents whether the relationship between X and Y is independent, positively correlated, or negatively correlated. Therefore, it can be used to determine the sign of the causal weight of the rule. Based on the association rules in Table 5.4, the causality matrix is derived from the causality weights between -1 and +1 as shown in Table 5.5.

Table 5.5 Result of causal matrix

		Market					Product						Technology				Outcome	
		Daily surveillance	Terrorism	Privacy regulation	Ecosystem disturbance	Political application	Logistics	Agricultural support	Disaster and safety	Warfare and weapons	Entertainment	Internet service	Platform and power technology	Detection avoidance	Navigation technology	Software technology	Economical value	Societal value
M	Daily surveillance																	−0.5
	Terrorism							−0.407	−0.376		−0.386		−0.376			−0.418		−0.9
	Privacy regulation	−0.441	−0.45									−0.434						0.6
	Ecosystem disturbance																	−0.5
	Political application	0.561		0.529														−0.5
P	Logistics	0.543			0.591												0.8	
	Agricultural support						0.644		0.635								0.4	0.3
	Disaster and safety												0.624				0.3	0.7
	Warfare and weapons		0.607								−0.425							0.3
	Entertainment	0.554					0.647										0.8	
	Internet service										0.575						0.7	0.5
T	Platform and power technology	0.568			0.617		0.633	0.639			0.649						0.4	0.4
	Detection avoidance				0.603											0.614	0.2	0.5
	Navigation technology						0.651			0.508	0.637	0.587		0.605			0.3	0.3
	Software technology	0.554				0.601		0.631		0.501			0.642				0.5	0.4
O	Economical value																	
	Societal value																	

5.4.4 Development of Weighted TRM

To construct the weighted TRM, the network, in which the 17 concepts correspond to nodes and the causal matrix corresponds to the links, was constructed. Like FCM, weights are expressed on nodes, causal weights are marked on links. The size of the node is proportional to the probability of occurrence. Also, the color and position of nodes are expressed according to the TRM layer classification of the concepts determined in Table 5.3. As a result, a weighted TRM was constructed as shown in Figure 5.7.

Figure 5.8 shows the result of evaluating the degree of centrality of each concept to understand the influence of concepts in the constructed FCM. First, platform and power technology and software technology concepts have very large outdegrees on other concepts. Developing platform and power technologies such as lightweight material technology, design automation technology, and high efficiency battery technology have the greatest influence on the next generation of UAV products and market. Also, software technologies such as improvement of UAV control and data processing have also played an important role in the development of UAV technology. On the other hand, the concept of daily surveillance concept showed the largest indegree than other concepts. As a result of technological developments and increasing use of UAVs, anyone will be able to surveillance and investigate, so that predictions and concerns about the occurrence of situations in which people are monitored daily coexist. In the case of entertainment and logistics concepts, outdegrees are bigger than indegrees, so total centralities are high. It is expected that the entertainment drones will have a large ripple effect not only for broadcasting shooting and production but also for personal use and leisure due to miniaturization and mobile phone linkage. Logistics UAVs are expected to develop rapidly in response to the commercialization of drone courier services by major global corporations, which will greatly affect the market and the environment.

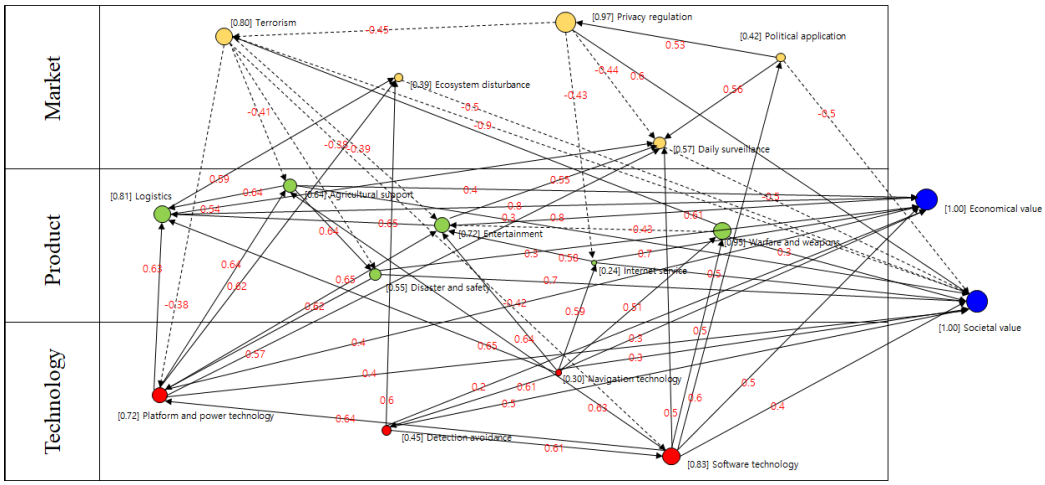

Figure 5.7 Result of weighted TRM development

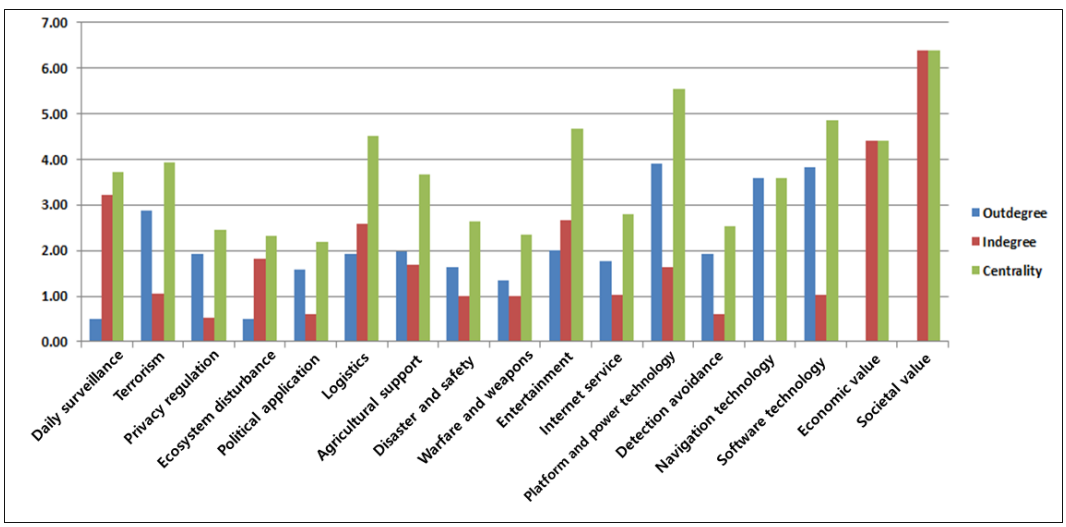

Figure 5.8 Centrality analysis of weighted TRM

5.4.5 Scenario Development

Through the above process, the various concepts and their relationship of the future of UAV technology are derived from the big data. However, it is unclear whether any of these concepts will necessarily occur and affect other concepts in the future. Therefore, it is

necessary to develop and analyze scenarios based on whether or not important concepts will occur. Based on the results of the centrality analysis, the concepts considered important were selected for morphological analysis. Among the concepts with high quantitative values, 8 concepts that are considered to be possible to develop more realistic and insightful scenarios were qualitatively chosen. The morphological matrix was constructed as shown in Table 5.6.

Table 5.6 Morphological matrix for scenario development

	Scenario projection			**Strategic options**				
	Daily surveillance	**Terrorism**	**Privacy regulation**	**Logistics**	**Entertainment**	**Platform and power technology**	**Detection avoidance**	**Software technology**
Change 1	Increased	Increased	Increased	Developed	Developed	Developed	Developed	Developed
Change 2	Not increased	Not increased	Lifted	Withdrawn	Withdrawn	Withdrawn	Withdrawn	Withdrawn

The significant input vector is set as follows after eliminated combinations of contradictory pairs in each concept dimension.

- Basic Scenario: All concepts occur
- Scenario 1: Lifting of privacy restrictions
- Scenario 2: Withdrawing the development of UAV technology for logistics
- Scenario 3: Failure of detection avoidance technology innovation

5.4.6 Dynamic Simulation of Scenarios

As shown in Table 5.7, the final output vector was derived by applying the dynamic reasoning mechanism of fuzzy awareness for each scenario. First, when the privacy regulation is lifted, the use of UAVs for terrorism (+0.07), daily surveillance activities (+0.06), political purposes (+0.06) and weapons (+0.03) increased significantly. Also due to concerns about privacy violation, development of UAVs for Internet service (-0.08), entertainment (-0.07) and logistics delivery (-0.05) is hampered and development of

navigation technology (-0.04) will be hindered. As a result, both the economic value and the social value decreased, especially the decline in social value is fatal. On the other hand, technologies and products related to agricultural support and emergency disaster, detection avoidance, software were found to be relatively independent.

Second, if the development of UAVs related to logistics is withdrawn, the development of navigation technology (-0.05) for UAV will be lowered, and development in the fields of agricultural support (-0.08) and disaster safety (-0.04) is expected to decrease. In terms of society and market, decreasing the use of UAVs reduces daily surveillance (-0.05) and ecosystem disturbance (-0.04). As a result, the economic value (-0.09) and the social value (-0.03) decreased, and the decline in economic value was relatively severe.

Third, if the innovations in technology for avoiding detection of UAVs fail, the ecosystem disturbance (+0.05) increases and the social value (-0.05) decreases. Software technology development (-0.07) also deteriorated, and overall development of UAV products such as logistics (-0.03), entertainment (-0.04), and internet service (-0.05) has been slightly lowered and economic value (-0.06) decline has occurred.

Table 5.7 Result of dynamic simulation

		Basic		**Scenario 1**		**Scenario 2**		**Scenario 3**	
		I	**O**	**I**	**O**	**I**	**O**	**I**	**O**
M	**Daily surveillance**	1	0.79	1	0.85	1	0.74	1	0.77
	Terrorism	1	0.52	1	0.59	1	0.51	1	0.53
	Privacy regulation	1	0.58	0	0	1	0.59	1	0.58
	Ecosystem disturbance	1	0.77	1	0.75	1	0.73	1	0.82
	Political application	1	0.58	1	0.64	1	0.58	1	0.56
P	**Logistics**	1	0.82	1	0.77	0	0	1	0.79
	Agricultural support	1	0.63	1	0.61	1	0.55	1	0.61
	Disaster and safety	1	0.55	1	0.55	1	0.51	1	0.54
	Warfare and weapons	1	0.63	1	0.66	1	0.63	1	0.62

	Entertainment	1	0.63	1	0.56	1	0.61	1	0.59
	Internet service	1	0.51	1	0.43	1	0.51	1	0.49
T	**Platform and power technology**	1	0.62	1	0.59	1	0.59	1	0.6
	Detection avoidance	1	0.58	1	0.58	1	0.57	0	0
	Navigation technology	1	0.5	1	0.46	1	0.45	1	0.5
	Software technology	1	0.53	1	0.52	1	0.53	1	0.46
O	**Economical value**	1	0.94	1	0.87	1	0.85	1	0.88
	Societal value	1	0.67	1	0.55	1	0.64	1	0.62
#of repetitions		21		24		22		20	

5.4.7 Final Scenario Roadmapping

Finally, a roadmap for each of the previously implemented multiple scenarios was constructed (See Figure 5.9-5.11). For the enterprise-level strategic roadmap, the decision making process according to the alternatives is additionally required. In this case study, the final roadmap is derived based on the dynamic simulation results, because it is aimed to create a predictive roadmap for the UAV industry.

Compared with the output vectors of the basic scenario, concepts which are enhanced or maintained in the output vector of each scenario were included in the corresponding scenario roadmap. In other words, the roadmaps drawn here focus only on the changes of causal relationships between the concepts derived as a result of the dynamic simulation for each scenario. Thus, some concepts may appear isolated or lack a link between other concepts on these roadmaps. For example, in the roadmap for scenario 2 (see Figure 5.10), the concept of “Entertainment” has only one connection with other concepts. Yet, this roadmap only shows the increased relationships in the scenario 2 that assumes the situation that withdraw the development of UAV technology for logistics. That is, the basic relationships between the “Entertainment” concept and other concepts are based on the weighted TRM in Figure 5.7, but are not represented in this roadmap.

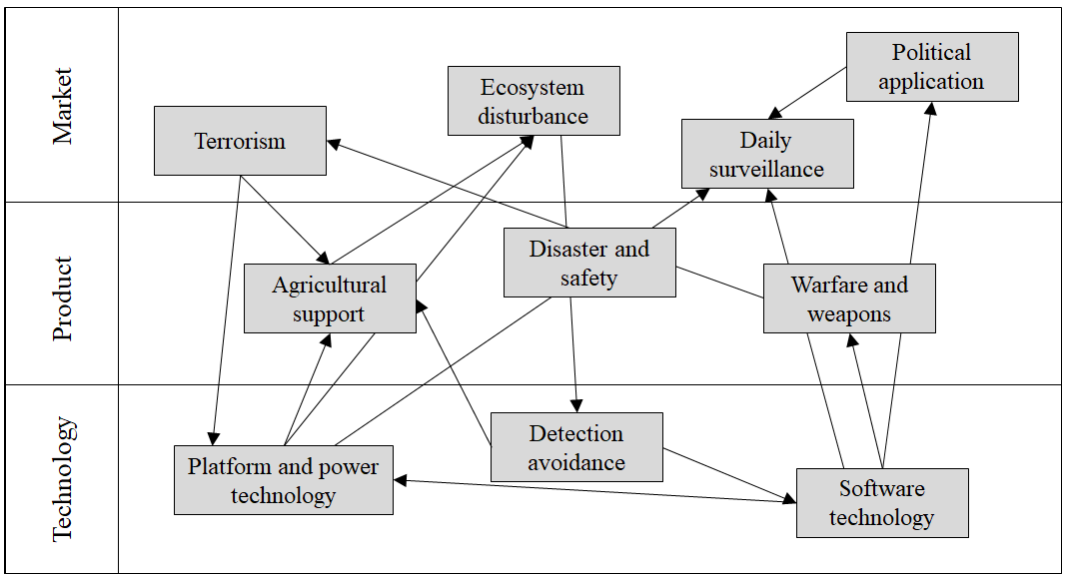

Figure 5.9 Roadmap for scenario 1 (Lifting of privacy restrictions)

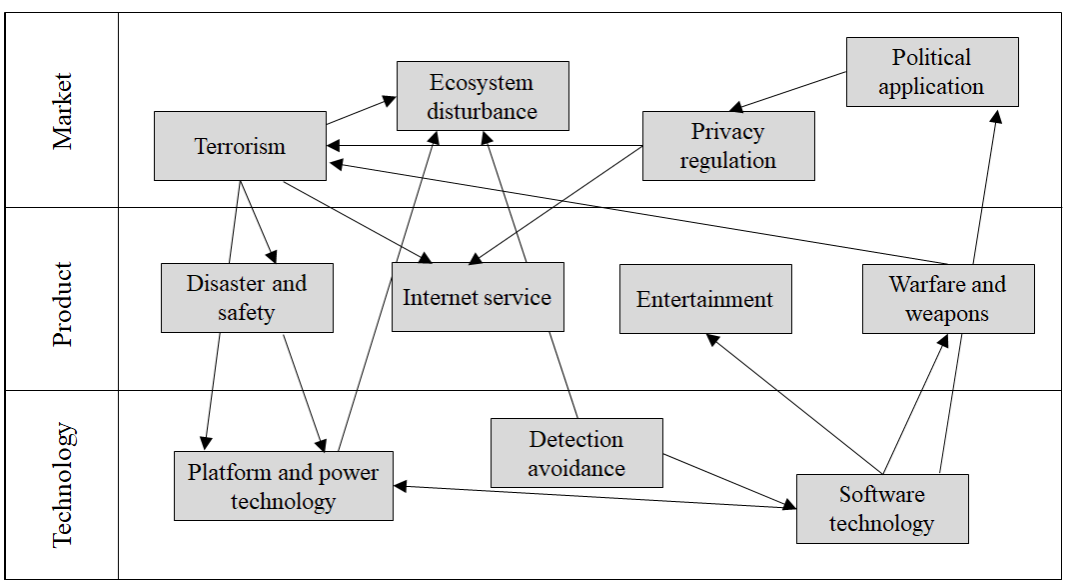

Figure 5.10 Roadmap for scenario 2 (Withdrawing the development of UAV technology for logistics)

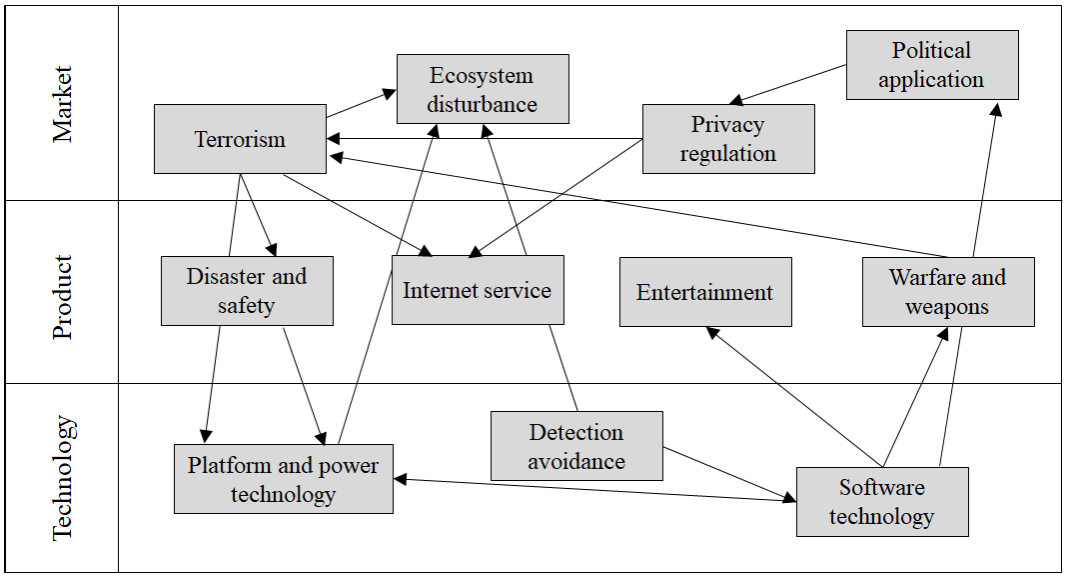

Figure 5.11 Roadmap for scenario 3 (Failure of detection avoidance technology innovation)

The implications from the scenario roadmap are as follows. First, the probability of occurrence of the concepts shown in each scenario roadmap is considered to be higher due to the simulation results. In other words, if a prerequisite situation arises in each scenario, the concepts are expected to be more important in the UAV industry. For example, the roadmap for Scenario 1 shows that the ten concepts in the roadmap have increased in importance. Second, it can be concluded that the linkage relationship in each roadmap also has a higher impact than in the basic scenario. Higher output vectors for each concept as a result of dynamic simulations indicate that their centralities also increased. In other words, as each concept's influence in the industry increases, the positive or negative impacts on other concepts may increase indirectly. Therefore, although the causality of the whole concepts is based on the connectivity of the basic scenario, it means that the connectivity shown in the final scenario roadmap is higher than that of the basic scenario.

5.5 Conclusion

This study examines the applicability of big data for emerging technology fields and suggests a big data application framework for scenario-based roadmapping. The research process consists of six steps that incorporate various qualitative and quantitative techniques. First, textual web data related to emerging technology is collected and the text mining is applied to identify the internal and external factors related to technology planning and occurrence probabilities. Second, ARM is applied to identify the cause-effect relationship between factors. Third, the network is visualized to construct the relationship between factors in the form of weighted TRM. Fourth, multiple scenario input vectors are generated based on the change of major factors using the scenario development method of FCM. Fifth, dynamic simulation for each scenario is carried out to obtain the final result vector of the scenario. Finally, multiple scenarios roadmapping is constructed based on scenario results. In order to confirm the usefulness of the framework proposed in this study, big data analysis for future scenario-based roadmapping of UAV technology was performed.

The proposed framework is expected to overcome some limitations of the existing scenario-based TRM, and the contributions can be summarized as follows. First, for the balanced consideration of internal and external uncertainties, the occurrence probability of factors corresponding to the external market, internal product and technology are all considered. In addition, a change in social and economic values are added to the scenario by adding an outcome layer. Second, various methodologies are integrated for the quantitative application of big data. Through this, various factors and correlations are extracted by grasping the semantic meaning from the unstructured textual data. Third, it is possible to construct various scenarios easily and flexibly through morphology analysis in the scenario development process. Therefore, this approach can effectively connect exploratory and divergent scenarios with purpose-oriented TRM.

However, there are some limitations in this study that require further improvement. The first and most crucial limitation is that time is not taken into account in the technology roadmap in the scenario development process presented in this study. One of the major

components of the technology roadmap is the time axis. Thus, the roadmap in the framework presented does not conform to the strict definition of technology roadmap. Rather, the term "issue-related roadmap" may be more appropriate than technology roadmap. However, the main goal of this study is to link the scenario planning in the technology planning stage with the roadmap, which is in line with the purpose of the technology roadmap. Overcoming these limitations requires coordination of the data collection steps. The data used in the case study of this research did not have a distinction of prediction time. Therefore, if data can be collected by dividing the prediction time into short-term, mid-term, and long-term periods, and a change pattern between factors over each period can be represented, a process can be developed that fits a more traditional technology roadmap. Third, more specific evaluation criteria are required for the scenarios derived from the process presented in this study. For example, in order to develop and select scenarios, occurrence probability, influence, genuineness, and usability of scenarios can be considered together. Second, the proposed methodology is mainly for the development planning and policy research of emerging technology, but the illustrative case study conducted in this study has established only an industry-level roadmap. Therefore, in future research, it is required to carry out an alternative evaluation of factors through cooperation with strategic planners and policy researchers who are users of roadmaps. Fourth, the quantitative approach through textual analysis is still lacking in evidence to be superior or alternatives to the insights of experts in each technology field. Consequently, big data-based scenario planning presented in this study will be able to deliver higher utility when utilized as a complementary to the results of the expert-based approach, and an additional framework for this will be required.

Chapter 6

Concluding Remarks

This thesis has aimed to build a public-based exploratory technology foresight framework. In order to overcome the limitations of existing expert-oriented normative future research, opinions of non-experts, or public were incorporated into the prediction process. By utilizing big data that exists online, collective intelligence for future prediction was realized. As a result, this study developed a research framework that can systematically conduct a sequential processes of screening, understanding, and analysis of big data.

As mentioned earlier, this study covered three research objectives: (1) To suggest criteria for evaluating and screening data sources for public-based technology foresight (2) To present a research framework that identify trends in public issues about new technologies (3) To develop a technology foresight process that derive future scenarios from big data.

The main contribution of this study is that it constructed a both qualitative and quantitative analysis process that collects, analyzes and exploit large amounts of unstructured text data to address these issues. This is expected to provide new insights in today's future research and contribute to expanding the scope of the forecasting process. Table 6.1 summarizes the theoretical and practical contributions of overall thesis and each research theme.

Table 6.1 Theoretical and practical contribution of this thesis

	Theoretical contribution	**Practical contribution**
Overall Thesis	- Proposes a systematic and subsequent research process for selecting, grasping and interpreting data - Contributes to broadening the horizon for technology foresight in the big data era	- Analyzes online text data to reflect public opinions and predictions on emerging technology both quantitatively and qualitatively
Chapter 3. Public Foresight Communities Evaluation	- Develops an online community evaluation for purpose of technology foresight - Combines bibliometric approaches and keyword-based text analysis	- Demonstrates that online community data is available as an alternative to expert-oriented technology forecasting
Chapter 4. Public Understanding of Technology	- Proposes a framework to quantitatively analyze the flow of changes in perception of technology through semantic analysis of large amounts of time series text data.	- Provides insight into how the public accepts and understands emerging technologies
Chapter 5. Scenario-based Foresight	- Suggests an approach that utilizes FCM and technology roadmap based on text data - Construct a data-driven quantitative scenario planning process	- Derive a predictive scenario through causal relationships among the elements that constitutes technology based on the actual public predictions

Chapter 3. Public Foresight Communities Evaluation

This study suggests a framework that evaluate the online community to select data sources for public-based technology foresight. Technology-related communities are assessed by two criteria: expertise and diversity. Specifically, text data relating to the future of emerging technologies was collected from online communities of various types. After that, four indicators were calculated ratio of author per document, optimal number of topics, level of knowledge, and ratio of activists. Based on these, the expertise and diversity scores were finally measured. For both criteria, online communities that meet or exceed a certain threshold were identified as ultimately a valuable data source. In order to verify the measurement method presented in this study, we applied the framework to 20 actual online communities.

The contribution of this study can be summarized as follows. From the theoretical perspective, first, the evaluation framework of the online community was newly proposed for the specific purpose of technology foresight. Traditional online community evaluation techniques focused on individual web pages and users rather than the community itself. In response, this study suggested a brand-new evaluation technique for the purpose of assessing data sources for technology foresight. This is expected to contribute to future research as new concept of open foresight emerges. Second, this study proposed a combined evaluation technique that utilizes a bibliometric approach and text analysis. A bibliometric approach based on the statistics of documents and authors within the online community allows indirect reflection of the size and user characteristics of the website. In addition, text analysis was further conducted to allow consideration of not only the external characteristics of the online community but also the contents contained therein. From the practical perspective, this study demonstrates that online community data is available as an alternative to expert-oriented technology forecasting. In today's big data era, it is important to verify the value of online community data as an alternative to costly expert data. The online data deemed valuable through this study is expected to provide new opportunities for future research.

Chapter 4. Public Understanding of Technology

This study suggests a framework that analyze interest of the public on emerging technologies. Textual web-based documents are utilized as a source of people's interest on emerging technologies. In detail, technology-related issues are extracted from online documents by making use of topic modeling technique, LDA. To explore the semantic changes of extracted topics over time, database is constructed by year and relationships between topics are measured based on similarity of their topical keywords. Then pathways that reflect change of public's interest on various technology fields were derived.

Contributions of this study can be summarized as follows. From the theoretical perspective, first, this research suggests a way to quantitatively measure the public's interest toward emerging technology. In this study, web-based textual documents, which is written by numerous people, are utilized rather than the scientific literatures or individual survey. Therefore, unlike previous studies which had difficulty in directly reflecting the opinions of the public, it was possible to analyze the perception of general users. Second, text analysis was performed over time through dynamic keyword-based approach. Traditional approaches of topic analysis are relatively static; they apply fixed models to the entire dataset and ignore any possible change resulting from time changes over time. In response, this study adopted time-series topic modeling technique, which is effective in estimating trends of topics extracted from document collection over time. In this way, it is possible to not only identify publicly focused issues but also explore the semantic change process simultaneously. From the practical perspective, this study provides insight into how the public accepts and understands emerging technologies. Through the data of people's opinions and predictions about various promising technologies, the trend of change in socio-technical perception was derived.

Chapter 5. Scenario-based Foresight

This study examines the applicability of big data for emerging technology fields and suggests a big data application framework for scenario-based roadmapping. The research process consists of six steps that incorporate various qualitative and quantitative techniques.

First, textual web data related to emerging technology is collected and the text mining is applied to identify the internal and external factors related to technology planning and occurrence probabilities. Second, ARM is applied to identify the cause-effect relationship between factors. Third, the network is visualized to construct the relationship between factors in the form of weighted TRM. Fourth, multiple scenario input vectors are generated based on the change of major factors using the scenario development method of FCM. Fifth, dynamic simulation for each scenario is carried out to obtain the final result vector of the scenario. Finally, multiple scenarios roadmapping is constructed based on scenario results. In order to confirm the usefulness of the framework proposed in this study, big data analysis for future scenario-based roadmapping of UAV technology was performed.

Contributions of this study can be summarized as follows. From the theoretical perspective, first, for the balanced consideration of internal and external uncertainties, the occurrence probability of factors corresponding to the external market, internal product and technology are all considered. In addition, a change in social and economic values are added to the scenario by adding an outcome layer. Second, various methodologies are integrated for the quantitative application of big data. Through this, various factors and correlations are extracted by grasping the semantic meaning from the unstructured textual data. From the practical perspective, it is possible to construct various scenarios easily and flexibly through causal relationships among the elements that constitutes technology based on the actual public predictions. Therefore, this approach can effectively connect exploratory and divergent scenarios through mixture use of FCM and TRM.

Despite these contributions, there still exists a substantial amount of work to be done for the development of a stronger theoretical and methodological foundation. Besides the limitations of each study that have been explained in the sub-conclusion of each chapter, the overall thesis has limitations as follows.

First, it is necessary to apply the framework proposed in this study to more technology fields. In the thesis, case studies related to a few specific emerging technologies were conducted, but it is required to analyze the various target technologies and prove the

effectiveness of the research. To this end, since data was collected from about 10 online communities in our research, it is needed to reconsider the applicability of numerous technology-related online communities.

Second, there is a need to develop more automated foresight processes. The development of today's promising technologies is very rapid, and convergence of technology is active. In order to effectively analyze these technologies, research process is needed to be done faster and with less human intervention. In other words, a framework that automatically collects and analyzes data and generates results is essential. Yet, the research process presented in this study still requires some qualitative intervention by the analyst. To overcome this, it would be helpful to adopt more advanced machine learning algorithms and novel IT such as XML for integrating various applications to analyze the data. In this case, more automated analysis is also possible, and the format of the data is expected to be more diversified not only with text data but also with images and videos.

Third, the prediction process and outcomes are needed to be diversified. In addition to the methodologies utilized in this thesis, online community data will be available for various approaches. For example, it may be possible to derive narrative scenarios through subject-action-object analysis of keywords, and to predict convergence between technologies through cross-impact analysis. If we apply the public's technological data to the methodologies that were previously focused on experts and those used for concept generation, a more colorful and original exploration of the future will be made.

Bibliography

[1] T. Albert, M.G. Moehrle and S. Meyer, Technology maturity assessment based on blog analysis, *Technological Forecasting and Social Change*, 92 (2015), pp. 196-209.

[2] N. Aletras and M. Stevenson, Measuring the similarity between automatically generated topics, in *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, 2014, pp. 22-27.

[3] M. Amer, T. Daim and A. Jetter, Application of fuzzy cognitive map for the development of scenarios: A case study of wind energy deployment, in *Research and Technology Management in the Electricity Industry*, Springer, LON, 2013, pp. 129-159

[4] H.J. An and S.J. Ahn, Emerging technologies—beyond the chasm: Assessing technological forecasting and its implication for innovation management in Korea, *Technological Forecasting and Social Change*, 102 (2016), pp. 132-142.

[5] R. Arun, V. Suresh, C.V. Madhavan and M.N. Murthy, On finding the natural number of topics with latent dirichlet allocation: Some observations, in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2010, pp. 391-402.

[6] R.U. Ayres, *Technological Forecasting and Long-range Planning*, McGraw-Hill, NY, 1969.

[7] A. Baram-Tsabari, M.W. Bauer, P. Jensen and E. Segev, Exploring new web-based tools to identify public interest in science, *Public Understanding of Science*, 20 (2009), pp. 130-143.

[8] V. Barassi and E. Treré, Does Web 3.0 come after Web 2.0? Deconstructing theoretical assumptions through practice, *New Media & Society*, 14 (2012), pp. 1269-1285.

[9] N. Beauchamp, Predicting and interpolating state-level polls using Twitter textual data, *American Journal of Political Science*, 61 (2017), pp. 490-503.

[10] F.M. Belz and W. Baumbach, Netnography as a method of lead user identification, *Creativity and Innovation Management*, 19 (2010), pp. 304-313.

[11] D.M. Blei and J.D. Lafferty, Dynamic topic models, in *Proceedings of the 23rd International Conference on Machine Learning*, 2006, pp. 113-120.

[12] D.M. Blei, A.Y. Ng and M.I. Jordan, Latent dirichlet allocation, *Journal of Machine Learning Research*, 3 (2003), pp. 993-1022.

[13] F. Bodendorf and C. Kaiser, Detecting opinion leaders and trends in online communities, in *2010 Fourth International Conference on Digital Society*, 2010, pp. 124-129.

[14] M. Borup, N. Brown, K. Konrad and H. Van Lente, The sociology of expectations in science and technology, *Technology Analysis & Strategic Management*, 18 (2006), pp. 285-298.

[15] P. Bosch-Sijtsema and J. Bosch, User involvement throughout the innovation process in high-tech industries, *Journal of Product Innovation Management*, 32 (2015), pp. 793-807.

[16] A. Brem and V. Bilgram, The search for innovative partners in co-creation: Identifying lead users in social media through netnography and crowdsourcing, *Journal of Engineering and Technology Management*, 37 (2015), pp. 40-51.

[17] J. Cao, T. Xia, J. Li, Y. Zhang and S. Tang, A density-based method for adaptive LDA model selection, *Neurocomputing*, 72 (2009), pp. 1775-1781.

[18] H. Chen, G. Zhang, D. Zhu and J. Lu, Topic-based technological forecasting based on patent data: A case study of Australian patents from 2000 to 2014, *Technological*

Forecasting and Social Change, 119 (2017), pp. 39-52.

[19] X. Chen and T. Han, Disruptive technology forecasting based on Gartner Hype Cycle, in *2019 IEEE Technology & Engineering Management Conference*, 2019, pp. 1-6.

[20] M. Cheng, J.W. Wong, C. Cheung and K. Leung, A scenario-based roadmapping method for strategic planning and forecasting: A case study in a testing, inspection and certification company, *Technological Forecasting and Social Change*, 111 (2016), pp. 44-62.

[21] Y. Cho and T. Daim, Technology forecasting methods, in *Research and Technology Management in the Electricity Industry,* Springer, LON, 2013, pp. 67-112.

[22] H. Çifci and N. Yüksel, Foresight 6.0: The new generation of technology foresight, in *2018 IEEE International Conference on Engineering, Technology and Innovation*, 2018, pp. 1-5.

[23] S.E. Cozzens, Distributive justice in science and technology policy, *Science and Public Policy*, 34 (2007), pp. 85-94.

[24] O. Da Costa, R. Cachia and R. Compañó, Can online social networks be used in forward-looking studies, in *Second International Seville Seminar on Future-Oriented Technology Analysis*, 2006, pp. 1-8.

[25] L. Dahlander, L. Frederiksen and F. Rullani, Online communities and open Innovation, *Industry and Innovation*, 15 (2009), pp. 115-123.

[26] T.U. Daim, G. Rueda, H. Martin and P. Gerdsri, Forecasting emerging technologies: Use of bibliometrics and patent analysis, *Technological Forecasting and Social Change*, 73 (2006), pp. 981-1012.

[27] O. Dedehayir and M. Steinert, The hype cycle model: A review and future directions, *Technological Forecasting and Social Change*, 108 (2016), pp. 28-41.

[28] R. Deveaud, E. SanJuan and P. Bellot, Accurate and effective latent concept modeling for ad hoc information retrieval, *Document Numérique*, 17 (2014), pp. 61-84.

[29] G. Di Stefano, A. Gambardella and G. Verona, Technology push and demand pull perspectives in innovation studies: Current findings and future research directions, *Research Policy*, 41 (2012), pp. 1283-1295.

[30] S.A. Drew, Building technology foresight: using scenarios to embrace innovation, *European Journal of Innovation Management*, 9 (2006), pp. 241-257.

[31] S.T. Dumais, Latent semantic analysis, *Annual Review of Information Science and Technology*, 38 (2004), pp. 188-230.

[32] D. Ehls and C. Herstatt, Open source participation behavior-A review and introduction of a participation lifecycle model, in *35th DRUID Celebration Conference*, 2013, pp. 1-26.

[33] J. Erik Karlsen and H. Karlsen, Expert groups as production units for shared knowledge in energy foresights, *Foresight*, 9 (2007), pp. 37-49.

[34] A.K. Firat, W.L. Woon and S. Madnick, Technological forecasting–A review, *Working Paper of Composite Information Systems Laboratory*, 2008, pp. 1-19.

[35] J.W. Forrester, Industrial dynamics, *Journal of the Operational Research Society*, 48 (1997), pp. 1037-1041.

[36] N. Gerdsri, An analytical approach to building a technology development envelope (TDE) for roadmapping of emerging technologies, *International Journal of Innovation and Technology Management*, 4 (2007), pp. 121-135.

[37] Y. Geum, H. Lee, Y. Lee and Y. Park, Development of data-driven technology roadmap considering dependency: An ARM-based technology roadmapping, *Technological Forecasting and Social Change*, 91 (2015), pp. 264-279.

[38] Y. Geum, S. Lee and Y. Park, Combining technology roadmap and system dynamics simulation to support scenario-planning: A case of car-sharing service, *Computers & Industrial Engineering*, 71 (2014), pp. 37-49.

[39] O. Glassey, Folksonomies: Spontaneous crowd sourcing with online early detection

potential?, *Futures*, 44 (2012), pp. 257-264.

[40] A. Gok, A. Waterworth and P. Shapira, Use of web mining in studying innovation, *Scientometrics*, 102 (2015), pp. 653-671.

[41] T.J. Gordon and H. Hayward, Initial experiments with the cross impact matrix method of forecasting, *Futures*, 1 (1968), pp. 100-116.

[42] T.L. Griffiths and M. Steyvers, Finding scientific topics, in *Proceedings of the National academy of Sciences*, 2004, pp. 5228-5235.

[43] I. Guy, U. Avraham, D. Carmel, S. Ur, M. Jacovi and I. Ronen, Mining expertise and interests from social media, in *Proceedings of the 22nd International Conference on World Wide Web*, 2013, pp. 515-526.

[44] K. Haegeman, E. Marinelli, F. Scapolo, A. Ricci and A. Sokolov, Quantitative and qualitative approaches in Future-oriented Technology Analysis (FTA): From combination to integration?, *Technological Forecasting and Social Change*, 80 (2013), pp. 386-397.

[45] I. Hagel and A.G. Armstrong, Net gain, *McKinsey Quarterly*, (1997), pp. 140-153.

[46] C. Hansen, T. Daim, H. Ernst and C. Herstatt, The future of rail automation: A scenario-based technology roadmap for the rail automation market, *Technological Forecasting and Social Change*, 110 (2016), pp. 196-212.

[47] J.C. Harper, Impact of technology foresight, *Nesta Working Paper*, 13(2013).

[48] T.H. Haveliwala, Topic-sensitive pagerank, in *Proceedings of the 11th International Conference on World Wide Web*, 2002, pp. 517-526.

[49] Q. He, B. Chen, J. Pei, B. Qiu, P. Mitra and L. Giles, Detecting topic evolution in scientific literature: how can citations help?, in *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, 2009, pp. 957-966.

[50] A. Heiko, V.A. Bañuls, M. Turoff, A.M. Skulimowski and T.J. Gordon, Foresight support systems: The future role of ICT for foresight, *Technological Forecasting and*

Social Change, 97 (2015), pp. 1-6.

[51] T. Hofmann, Probabilistic latent semantic indexing, in *ACM SIGIR Forum*, 2017, pp. 211-218.

[52] L. Hong and B.D. Davison, Empirical study of topic modeling in twitter, in *Proceedings of the First Workshop on Social Media Analytics*, 2010, pp. 80-88.

[53] I.M. Ilevbare, D. Probert and R. Phaal, Towards risk-aware roadmapping: Influencing factors and practical measures, *Technovation*, 34 (2014), pp. 399-409.

[54] L. Janzik and C. Raasch, Online communities in mature markets: Why join, why innovate, why Share?, *International Journal of Innovation Management*, 15 (2011), pp. 797-836.

[55] A. Jetter and W. Schweinfort, Building scenarios with Fuzzy Cognitive Maps: An exploratory study of solar energy, *Futures*, 43 (2011), pp. 52-66.

[56] V. Kayser, Comparing public and scientific discourse in the context of innovation systems, *Technological Forecasting and Social Change*, 115 (2017), pp. 348-357.

[57] D. Kim and A. Oh, Topic chains for understanding a news corpus, in *International Conference on Intelligent Text Processing and Computational Linguistics*, 2011, pp. 163-176.

[58] J. Kim, Y. Park and Y. Lee, A visual scanning of potential disruptive signals for technology roadmapping: investigating keyword cluster, intensity, and relationship in futuristic data, *Technology Analysis & Strategic Management*, 28 (2016), pp. 1225-1246.

[59] K. Klincewicz, The emergent dynamics of a technological research topic: the case of graphene, *Scientometrics*, 106 (2016), pp. 319-345.

[60] A.R. Köhler and C. Som, Risk preventative innovation strategies for emerging technologies the cases of nano-textiles and smart textiles, *Technovation*, 34 (2014), pp. 420-430.

[61] K. Kok, The potential of Fuzzy Cognitive Maps for semi-quantitative scenario development, with an example from Brazil, *Global Environmental Change*, 19 (2009), pp. 122-133.

[62] B. Kosko, Fuzzy cognitive maps, *International Journal of Man-machine Studies*, 24 (1986), pp. 65-75.

[63] G. Krampen, A. Von Eye and G. Schui, Forecasting trends of development of psychology from a bibliometric perspective, *Scientometrics*, 87 (2011), pp. 687-694.

[64] H. Kwon, J. Kim and Y. Park, Applying LSA text mining technique in envisioning social impacts of emerging technologies: The case of drone technology, *Technovation*, 60 (2017), pp. 15-28.

[65] J. Lazar and J. Preece, Classification schema for online communities, in *Proceedings of the Association for Information Systems 1998 Americas Conference*, 1998, pp. 84-86.

[66] U. Lechner, K. Stanoevska-Slabeva and Y.-H. Tan, Communities in the digital economy, *International Journal of Electronic Commerce*, 6 (2002), pp. 5-7.

[67] C. Lee, B. Song and Y. Park, An instrument for scenario-based technology roadmapping: How to assess the impacts of future changes on organisational plans, *Technological Forecasting and Social Change*, 90 (2015), pp. 285-301.

[68] C.J. Lee, D.A. Scheufele and B.V. Lewenstein, Public attitudes toward emerging technologies: Examining the interactive effects of cognitions and affect on public attitudes toward nanotechnology, *Science Communication*, 27 (2005), pp. 240-267.

[69] H.J. Lee, S. Lee and B. Yoon, Technology clustering based on evolutionary patterns: The case of information and communications technologies, *Technological Forecasting and Social Change*, 78 (2011), pp. 953-967.

[70] L. Liu, L. Tang, W. Dong, S. Yao and W. Zhou, An overview of topic modeling and its current applications in bioinformatics, *SpringerPlus*, 5 (2016), p. 1608.

[71] H. Lu and H. You, Roadmap modeling and assessment approach for defense technology system of systems, *Applied Sciences*, 8 (2018), p. 908.

[72] W. Luo, B. Stenger, X. Zhao and T.K. Kim, Automatic topic discovery for multi-object tracking, in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015, pp. 3820-3826.

[73] H.R. Maier, J.H. Guillaume, H. van Delden, G.A. Riddell, M. Haasnoot and J.H. Kwakkel, An uncertain future, deep uncertainty, scenarios, robustness and adaptation: How do they fit together?, *Environmental Modelling & Software*, 81 (2016), pp. 154-164.

[74] J.P. Martino, *Technological Forecasting for Decision Making*, McGraw-Hill, NY, 1993.

[75] E.B. Masini, The past and the possible futures of Futures Studies: Some thoughts on Ziauddin Sardar's 'the namesake', *Futures*, 42 (2010), pp. 185-189.

[76] A. McAfee, E. Brynjolfsson, T.H. Davenport, D. Patil and D. Barton, Big data: the management revolution, *Harvard Business Review*, 90 (2012), pp. 60-68.

[77] V. Miemis, J. Smart and A. Brigis, Open foresight, *Journal of Futures Studies*, 17 (2012), pp. 91-98.

[78] M. Nisbet and E.M. Markowitz, Understanding public opinion in debates over biomedical research: looking beyond political partisanship to focus on beliefs about science and society, *PloS one*, 9 (2014), p. e88473.

[79] S.B. Nolen, The role of literate communities in the development of children's interest in writing, in *Writing and Motivation*, Brill, UK, 2006, pp. 239-255

[80] K. Okamura, Dynamic development of public attitudes towards science policymaking, *Public Understanding of Science*, 25 (2016), pp. 465-479.

[81] T. O'reilly, What is Web 2.0: Design patterns and business models for the next generation of software, *International Journal of Digital Economics*, 65 (2007), pp. 17-

37.

[82] M. Pagani, Roadmapping 3G mobile TV: Strategic thinking and scenario planning through repeated cross-impact handling, *Technological Forecasting and Social Change*, 76 (2009), pp. 382-395.

[83] D.H. Palmer, Student interest generated during an inquiry skills lesson, *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 46 (2009), pp. 147-165.

[84] R. Plant, Online communities, *Technology in Society*, 26 (2004), pp. 51-65.

[85] R. Popper, Foresight methodology, in *The Handbook of Technology Foresight*, Edward Elgar, UK, 2008, pp. 44-88

[86] J. Preece, Sociability and usability in online communities: Determining and measuring success, *Behaviour & Information Technology*, 20 (2010), pp. 347-356.

[87] J. Rech, Discovering trends in software engineering with google trend, *ACM SIGSOFT Software Engineering Notes*, 32 (2007), pp. 1-2.

[88] K.A. Renninger and S. Hidi, Revisiting the conceptualization, measurement, and generation of interest, *Educational Psychologist*, 46 (2011), pp. 168-184.

[89] A. Retzbach, J. Marschall, M. Rahnke, L. Otto and M. Maier, Public understanding of science and the perception of nanotechnology: the roles of interest in science, methodological knowledge, epistemological beliefs, and beliefs about science, *Journal of Nanoparticle Research*, 13 (2011), pp. 6231-6244.

[90] J. Retzbach, L. Otto and M. Maier, Measuring the perceived uncertainty of scientific evidence and its relationship to engagement with science, *Public Understanding of Science*, 25 (2016), pp. 638-655.

[91] O. Saritas and J. Aylen, Using scenarios for roadmapping: The case of clean production, *Technological Forecasting and Social Change*, 77 (2010), pp. 1061-1075.

[92] O. Saritas, L.A. Pace and S.I. Stalpers, Stakeholder participation and dialogue in

foresight, in *Participation and Interaction in Foresight: Dialogue, Dissemination and Visions*, Edward Elgar, UK, 2013, pp. 35-69

[93] J. Schatzmann, R. Schäfer and F. Eichelbaum, Foresight 2.0-Definition, overview & evaluation, *European Journal of Futures Research*, 1 (2013), pp. 1-15.

[94] D.A. Scheufele and B.V. Lewenstein, The public and nanotechnology: How citizens make sense of emerging technologies, *Journal of Nanoparticle Research*, 7 (2005), pp. 659-667.

[95] R. Siebelink, J.I. Halman and E. Hofman, Scenario-driven roadmapping to cope with uncertainty: Its application in the construction industry, *Technological Forecasting and Social Change*, 110 (2016), pp. 226-238.

[96] P.J. Silvia, Interest and interests: The psychology of constructive capriciousness, *Review of General Psychology*, 5 (2001), pp. 270-290.

[97] J.M. Sims, Communities of practice: Telemedicine and online medical communities, *Technological Forecasting and Social Change*, 126 (2018), pp. 53-63.

[98] C.O. Stewart, D.L. Dickerson and R. Hotchkiss, Beliefs about science and news frames in audience evaluations of embryonic and adult stem cell research, *Science Communication*, 30 (2009), pp. 427-452.

[99] J.D. Strauss and M. Radnor, Roadmapping for dynamic and uncertain environments, *Research-Technology Management*, 47 (2004), pp. 51-58.

[100] M. Sturken, D. Thomas and S. Ball-Rokeach, *Technological Visions: The Hopes and Fears that Shape New Technologies*, Temple University Press, PA, 2004.

[101] A. Suominen and H. Toivanen, Map of science with topic modeling: Comparison of unsupervised learning and human-assigned subject classification, *Journal of the Association for Information Science and Technology*, 67 (2016), pp. 2464-2476.

[102] B. Takahashi and E.C. Tandoc Jr., Media sources, credibility, and perceptions of science: Learning about how people learn about science, *Public Understanding of*

Science, 25 (2016), pp. 674-690.

[103] Y.W. Teh, M.I. Jordan, M.J. Beal and D.M. Blei, Sharing clusters among related groups: Hierarchical Dirichlet processes, in *Advances in Neural Information Processing Systems*, 2005, pp. 1385-1392.

[104] H. Van Lente, C. Spitters and A. Peine, Comparing technological hype cycles: Towards a theory, *Technological Forecasting and Social Change*, 80 (2013), pp. 1615-1628.

[105] C. Wang, D. Blei and D. Heckerman, Continuous time dynamic topic models, in *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence*, 2012, pp. 579-586.

[106] G.A. Wang, J. Jiao, A.S. Abrahams, W. Fan and Z. Zhang, ExpertRank: A topic-aware expert finding algorithm for online knowledge communities, *Decision Support Systems*, 54 (2013), pp. 1442-1451.

[107] S. Whittaker, E. Isaacs and V. O'Day, Widening the net: workshop report on the theory and practice of physical and network communities, *SIGCHI Bulletin*, 29 (1997), pp. 27-30.

[108] D. Xenias and L. Whitmarsh, Dimensions and determinants of expert and public attitudes to sustainable transport policies and technologies, *Transportation Research Part A: Policy and Practice*, 48 (2013), pp. 75-85.

[109] C.K. Yau, A. Porter, N. Newman and A. Suominen, Clustering scientific documents with topic modeling, *Scientometrics*, 100 (2014), pp. 767-786.

[110] C.A. Yeung, M.G. Noll, N. Gibbins, C. Meinel and N. Shadbolt, SPEAR: spamming-resistant expertise analysis and ranking in collaborative tagging systems, *Computational Intelligence*, 27 (2011), pp. 458-488.

[111] M.A. Zeng, The contribution of different online communities in open innovation projects, in *Proceedings of The International Symposium on Open Collaboration*,

2014, pp. 1-9.

[112] M.A. Zeng, Foresight by online communities – The case of renewable energies, *Technological Forecasting and Social Change*, 129 (2018), pp. 27-42.

[113] J. Zhang, M.S. Ackerman and L. Adamic, Expertise networks in online communities: structure and algorithms, in *Proceedings of the 16th International Conference on World Wide Web*, 2007, pp. 221-230.

[114] Y. Zhang, H. Chen, J. Lu and G. Zhang, Detecting and predicting the topic change of Knowledge-based Systems: A topic-based bibliometric analysis from 1991 to 2016, *Knowledge-Based Systems*, 133 (2017), pp. 255-268.

[115] Y. Zhang, D.K. Robinson, A.L. Porter, D. Zhu, G. Zhang and J. Lu, Technology roadmapping for competitive technical intelligence, *Technological Forecasting and Social Change*, 110 (2016), pp. 175-186.

[116] Y. Zhang, G. Zhang, D. Zhu and J. Lu, Scientific evolutionary pathways: Identifying and visualizing relationships for scientific topics, *Journal of the Association for Information Science and Technology*, 68 (2017), pp. 1925-1939.

Appendix

Appendix A Results of topic modeling by year

2011

Topic	Nano medicine	Robot regulation and social influence	Modern warfare	Tablet devices and ebooks	Advance into space	Social media and entertainment	High definition video technology	Climate change and global warming
Keywords	genetic cells cell nanobots body gene replace hair eyes skin bacteria nanotubes	robots laws government online economy material information price cost break jobs law hope system society	living inhabitants military battlefield weapons indistinguishable environment war terrain unstable	mobile market kindle devices tablet computer tablets screen lcd apple books amazon ereader ipad	world mars earth faster asteroid space moon planet living	tv time games facebook movie glasses social media speed online experience free home	hd blu tv sense increase display pixels surface efficient resolution screen rays grow	climate change earth energy warming global sun weather nature impact atmosphere

2012

Topic	Mutant species	Emergence of VR	Brain science	Child education	Stem cell medicine	Technology regulation and media	Automation society
Keywords	species alien imagine bad detection earth space planet predatory animals aggressive	virtual world reality real ram gaming infinite feel idea travel	brain cleverbot stem user information surgery faster regeneration synthetic nanobots treatment	brain mind learning experience easier knowledge talking children play feel ability continuity	cells stem energy meat cancer cell gene medicine printing organs skin weight disease	technology support commercials society impede political military government free nuclear capitalism politics ideologies advancement news	computer robots drive farming warp flying advanced speed stress life economic population cities

2013

Topic	Robotic medicine and longevity	VR applications	Space travel	Rare species extinction	Contemporary social dispute	Robots and AI	3D printing technology	Space physics
Keywords	robots surgery body medical hope nanotechnology nanobots brain dna genome cells docs longevity	virtual reality tv games internet watch computer smart phone screen augmented oculus	universe infinite travel gravity space earth reality station interstellar	life species real animals die extinct food century death bionic waste dolphins	world war teleportation happen money human aging nuclear robots countries drugs power weapons eventually china	computers singularity robots intelligence ai brain mind quantum information	3d printer technology printing materials electric clothing speed build house manufacturing generation	mars earth space moon solar future orbit planets mercury nasa

2014

Topic	Extraterrestrial life	AI and machine learning	Medical biology	Quantum physics	Boundary between VR and reality	Internet of things	Electric vehicle
Keywords	earth species planets universe planet alien stars star solar aliens sun galaxy travel asteroids	ai machine learning algorithms neural language computer google deep memory watson jobs network robots	medical life cure immune blood meat nanobots disease organs heart medicine tissue pain	quantum matter theory ftl particles discover warp physics entanglement	real die vr technology simulation virtual body born god imagine death family gender games forever love	internet technology world google smart games chip tv control school phone house education	energy power cars carbon water cost electric tesla gas batteries fuel expensive

2015

Topic	Space exploration	AI development	Elderly care	VR technology	Alternative energy	Nanotechnology
Keywords	earth space moon solar mars planets population resources travel eventually mission civilization probe aliens ftl	ai brain intelligence computers idea future robots imagine thinking entropy	people hair anti aging psychology mental pets drugs mind thinking therapy learning	vr real future virtual move reality physical tv holographic medical heart	time power water solar cars food government terraforming oxygen energy radiation low electric surface	red quark chamber gluon anti nanotech physics quantum bioterrorism

2016

Topic	The Singularity of AI	Immortality	Robotics Ethics	Resource depletion	Survival of mankind	Space exploration	Solar system
Keywords	ai intelligence singularity happen computers learning sentience artificial neural advanced close consciousness deepmind quantum	body life time energy immortality power aging forever memories hair chance virtual sleep simulation die genetic	robots system evil evolution humanity species humanoid aliens clone wrong	energy light drew power heat matter earth dark solar system law source nuclear	people time life population food water control global cancer exist disease hope bad	universe future space ship wormhole travel speed experience weapon craft drones warming change	space mars moon earth solar nasa asteroid missions spacex launch travel gravity send

2017

Topic	Climate change and global warming	Deep learning age	Diffusion of VR	Demographic change	Future transportation	Solar system travel	Posthuman	New energy and crisis
Keywords	climate change global geoengineering level rise earth sea ice warming heat water ocean hurricanes temperature atmosphere	ai brain human neural learning computers computer deep networks intelligence quantum network games faster	vr reality real virtual dimensional human screen imagine space body phones humanoid hardware home displays ar image	people sex human trillion women world clone single society change post memory school	cars space flying energy car technology cities material building black driverless countries speed vantablack	space earth mars solar planet star moon nuclear atmosphere interstellar light jupiter travel antimatter colonization gravity iss	life time human aging future physics anti born civilization posthuman humanity body living cancer gene cells diseases	energy fusion nuclear power solar fuel reactors war atoms hydrogen renewable

국문초록

오늘날, 기술 예측 과정에서의 대중 의견 반영에 대한 필요성이 지속적으로 증가하고 있다. 유망 기술의 확산으로 인해 미래에 발생할 수 있는 사회적 영향 및 위험에 대한 인식과 관점의 측면에서 전문가와 대중 간 차이가 존재할 수 있기 때문이다. 한편, 최근 정보통신 기술과 온라인 커뮤니티의 급격한 발전을 통해 과거에는 접근이 힘들었던 대량의 정보와 지식이 공유되고 있다. 이를 통해 다양한 분야의 사람들의 기술에 대한 의견을 포함한 웹 데이터가 축적되고 있으며 이러한 변화는 미래 연구를 위해 전통적으로 이루어지던 정성적인 연구를 벗어난 새로운 기술 예측의 기회를 제공하고 있다. 이에 대응하여, 본 논문은 기존의 규범적 기술 예측과는 달리 더욱 다양한 미래의 가능성에 대해 분석하는 대중 기반의 탐색적 기술 예측을 위한 새로운 프레임워크를 제시하고자 한다. 이를 위해, 다음의 세 가지 목적에 대하여 각각의 연구주제를 통해 분석을 진행한다. (1) 기술예측을 위한 데이터 소스의 평가 및 선별 프로세스를 제시한다 (2) 새로운 기술에 대한 대중들의 주요 관심 이슈의 변화 흐름을 파악할 수 있는 분석 프레임워크를 제시한다 (3) 온라인 커뮤니티 데이터를 통해 유망 기술에 대한 대중들의 예측 시나리오를 도출하기 위한 방법론을 개발한다.

첫번째 연구는 수많은 온라인 커뮤니티 중, 어떤 데이터가 기술 예측에 적합한지 평가 및 선별하는 것을 목표로 한다. 이를 위해, 전문성과 다양성이라는

기준에 대해 온라인 커뮤니티에서 발생된 기술 예측 데이터를 평가하는 방법을 제시한다. 구체적으로는, 커뮤니티 내의 게시글에 관한 계량서지학적 접근과 게시글 내에 담긴 내용을 의미론적으로 분석하는 텍스트 마이닝 기법을 혼합하여 기술 지식 수준, 리드유저의 비율, 다뤄지는 주제의 다양성 등을 계산한다. 이를 토대로 최종적으로 기술 예측을 위한 적합한 데이터 원천으로 판단되는 온라인 커뮤니티를 선별한다.

두번째 연구는 대중들의 유망기술에 대한 인식의 변화 흐름을 분석하는 것을 목표로 한다. 즉, 앞서 수집한 데이터에 대한 심도 있는 이해를 위해 대중들과 유망기술의 관계를 심층적으로 분석하는 것이다. 이를 위해, 장기간에 걸쳐 누적된 대중들의 기술 관련 데이터에 대해 시간의 흐름에 따른 이슈의 변화를 분석하는데 유용한 동적 토픽모델링 기법을 적용한다. 결과적으로 다양한 주제들을 다루고 있는 온라인 커뮤니티 데이터로부터 특정 기술 분야와 관련된 토픽의 변화를 추적하고, 사회적 인식의 변화를 확인할 수 있다.

세번째 연구는 텍스트 분석을 통해 빅데이터로부터 유망기술의 미래와 관련된 요소들간의 관계를 파악하고, 이를 토대로 시나리오 기반 기술로드맵을 개발하는 것을 목표로 한다. 이를 위해, 시스템을 구성하는 개념들간의 인과관계를 확률적으로 나타내는 유향성 그래프인 퍼지인식도를 활용한 시나리오 분석을 시행한다. 결과적으로 비전문가로부터 발생된 데이터를 통해, 기획 단계에서 유망기술의 미래에 대한 정량적 예측을 도출하며, 이는 대중 기반의 기술 탐색에 대한 가능성을 제시한다.

본 논문은 온라인 커뮤니티 데이터의 선별, 이해, 해석이라는 일련의 과정을 통해 오늘날의 빅 데이터 시대에 필요한 새로운 기술 예측 및 탐색

프레임워크를 제시한다. 비정형 텍스트 데이터의 분석을 위해 키워드 기반의 텍스트 분석과 퍼지인식도를 활용한 시나리오 기획을 진행하며, 이를 통해 보다 탐색적인 동시에 정량적인 결과를 도출 가능할 것으로 기대된다. 기술의 미래에 대해 다양한 가능성을 확인할 수 있는 대중 기반의 기술 예측 프레임워크를 통해 기존 전문가 중심의 규범적 예측 활동과 상호보완적 영향관계가 이루어질 수 있을 것으로 예상한다.

주요어: 대중 기반 기술 예측, 온라인 커뮤니티 데이터, 텍스트 마이닝, 토픽 모델링, 시나리오 기획

학번: 2014-21814