공학박사학위논문

Development and Utilization of GTM-based Maps using Information Visualization

정보시각화를 이용한

GTM 기반 지도의 개발 및 활용

2012년 8월

서울대학교 대학원

산업공학과

손 창 호

Development and Utilization of GTM-based Maps using Information Visualization

지도교수 박 용 태

이 논문을 공학박사 학위논문으로 제출함
2012년 5월

서울대학교 대학원
산업공학과
손 창 호

손창호의 공학박사 학위논문을 인준함
2012년 6월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

위 원 _____ (인)

Abstract

Development and Utilization of GTM-based maps using information visualization

Changho Son

Department of Industrial Engineering

The Graduate School

Seoul National University

Implementing information visualization has various advantages when to analyze voluminous information since the primary objective of information visualization is to construct a process that describes and explores specific information through graphical representation. That is, visualization amplifies human's cognitive ability and reduces the complex cognitive work necessary to perform certain activities. As well, unexplored insights are able to be provided through showing big pictures. Accordingly, this doctoral dissertation proposes a systematic approach to development and application of GTM-based maps using information visualization to identify new opportunities and explore trends or changes in technology or service. The dissertation is composed of three studies each of which addresses each of the three subjects: 1) systematic approach to identifying new technology opportunities through the generative topographic mapping (GTM) based patent vacuum map, 2) approach to identifying new service opportunities through the GTM-based service vacuum map, and 3) approach to analyzing service trends by the generative topographic mapping through time (GTM-TT) based service trend map. Text mining techniques are employed to transform unstructured textual items into structured data by using the vector space model for extracting and

analyzing valuable information from voluminous textual data. The GTM is a probabilistic model to mapping multidimensional data space onto a low-dimensional latent space and vice versa and provide to be a creditable alternative to the self organizing map (SOM) in terms of using a probabilistic method based on Bayesian theory. The GTM-TT is one such extension of GTM for the exploratory analysis of multivariate time series by performing simultaneous time series clustering and visualization.

In this research, discovering patent vacuums to identify new technology opportunities using GTM-based patent vacuum map is conducted. The patent map has long been considered as a useful tool for mining latent technological information. Among others, the detection of patent vacuums, defined as unexplored areas of new technologies, deserves intensive research. However, previous studies for identifying patent vacuums on the patent map have been subjected to some limitations, stemming from the subjective and manual identification of patent vacuums. To address these limitations, this study proposes a GTM-based patent vacuum map, which aims to automatically identify a patent vacuum. Since GTM is a probabilistic approach of mapping multidimensional data space onto a low-dimensional latent space and vice versa, it contributes to the automatic detection and interpretation of patent vacuums. The proposed approach consists of three stages. Firstly, text mining is executed in order to transform patent documents into keyword vectors as structured data. Secondly, the GTM is employed to develop the patent map, subsequently leading to the discovery of patent vacuums, which are expressed as blank areas in the map. Lastly, the meaning of each patent vacuum is interpreted as new technology opportunities by the inverse mapping of patent vacuums onto the original keyword vector. The case study is conducted with lithography technology-related patents. We believe the proposed approach not only saves time and effort for identifying patent vacuums, but also increases objectivity and reliability.

Unlike the above first study deals with technology area, second study concerned with identifying new service opportunities through derived service vacuums using GTM-based service vacuum map. Despite the strategic and technological gravity of new service opportunities, relatively little research has been devoted to the intelligent exploration and systematic identification of

new service opportunities. This study proposes a unique approach for developing and utilizing GTM-based service vacuum maps to discover new service opportunities. The detailed procedure of the approach is illustrated for the case of mobile application services from Apple's AppStore, which is a web service that allows smartphone users to access mobile application services. The proposed approach is expected to aid the discovery of new service opportunities from various information systems.

Lastly, identifying trends of service using GTM-TT service trend map is dealt with. Recently, due to the explosive increase of services, firms have faced with challenges to analyze patterns and trends in services in an intuitive but objective ways. The notion of service map can be adapted to this end. Maps, in general, have been receiving a great deal of attention because of their potential as visualization tools that can allow people to visualize massive amounts of information. Specifically, the GTM-TT algorithm is suitable for dynamic analysis since GTM-TT provides a time-based clustering and change path. In response, this study proposes an approach for developing and using GTM-TT service trend maps consisting of a service clustering map and a service sequence map for analyzing service trends. The proposed approach, broadly, is comprised of four steps: 1) the construction of a database, 2) data preprocessing, 3) development of a GTM-TT service trend map, and 4) interpretation. The proposed approach is expected to aid in the identification of dynamic service trends for other service areas as well.

Keywords: Information visualization, Text mining, generative topographic mapping (GTM), Generative topographic mapping through time (GTM-TT).
Student Number: 2009-30757

Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Background and motivations

Technology is one of the most important elements for providing companies with outstanding revenue in the current competitive environment. Hence, companies operating in competitive environments demanding new technology and product development, process improvement, and technology-enhanced services must achieve and organize information on emerging technologies (Choi and Park, 2009). As well, as a consequence of the service economy, R&D of the service industry has become more essential nowadays (Suh and Park, 2009). Therefore, the importance of service innovation for improving service firm's competitiveness is increasing with the phenomenon such as the deregulation and globalization of markets, the emergence of heightened competition, increased heterogeneity of customer demands, shortened product life cycles, and the internationalization of service firms (Eric and Sergios, 2005; Menor and Roth, 2008). In terms of technology management, only technology by research such as R&D management, technology forecasting, and so on had been dealt with. However, services, especially, technology-based services have been considered along with technology as application areas of technology management.

In this circumstance, the strategic importance of identifying new opportunities and understanding changes in technology and service for successful business of most firms increases. Furthermore, the ability to explore undeveloped areas of technology and service and monitor the current stage and history of technology and service is reckoned as a critical asset both for gaining competitive advantage and identifying promising niches (Choi and Park, 2009).

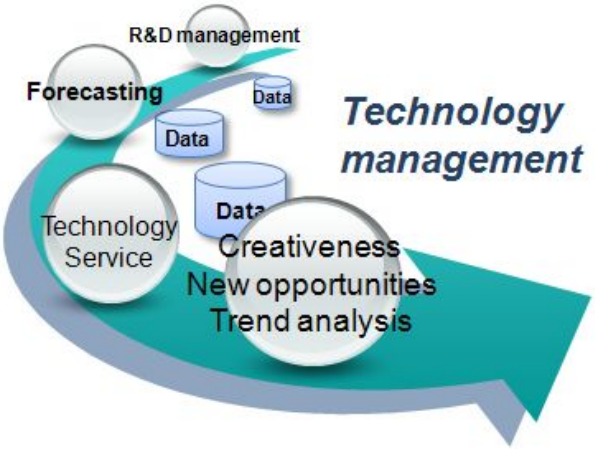


Figure 1-1 The trend of technology management

Furthermore, the trend of technology management has been changed. In the past, R&D management and forecasting for only technology were focused. However, recently, both technology and service have been studied in technology management field as shown in Figure 1-1. In contrast to the past decades, there are already plenty of data directly or indirectly related to technology or service and more data will be generated in the future as the

importance of technology and service is increased. Accordingly, much effort has been made to develop frameworks, models, tools, techniques, and methodologies to aid identifying new opportunities and analyzing trends in technology or service using data-driven approaches to overcome drawbacks of previous studies relying on experts' opinions and manual works. Nonetheless, there is a critical limitation, which is a difficulty of common designers of technology or service to understand the results, in practice, since the results of previous studies have been mainly derived from statistical or complex analysis.

In response, the concept of information visualization is used in this dissertation since this concept is very useful to analyze and interpret the results. Information visualization is defined as the use of computer supported, interactive, visual representations of data to amplify cognition (Card et al., 1999). Visual analytics reduces the user's cognitive burden by combining and leveraging both human and electronic data processing strengths and capabilities. The human visual system provides with the capability to quickly identify patterns and structures (van Wijk, 2005) and supports the transition from cognition, the processing of information, to perception, the obtaining of insight and knowledge. Hence, visual representations are often the preferred form of support to any human cognitive task because they amplify human's cognitive ability (Russell et al., 2009) and reduce the complex cognitive work necessary to perform certain activities (Keim, 2001; Keim et al., 2008). Information visualization is essential to scientific reasoning and the scientific process as a whole. However, information visualization should be considered

as an exploration means and not limited to communication or presentation purposes (van Wijk, 2005). In other words, to facilitate the generation of knowledge and the formulation of informed decisions, information visualization needs to be combined with analytical techniques (Keim et al., 2008) and embedded in the analysis/reasoning process (Meyer et al., 2007).

This dissertation proposed GTM-based maps, which are defined as two-dimensional maps using information visualization techniques such as generative topographic mapping (GTM) for identifying vacuums in technology or service, and generative topographic mapping through time (GTM-TT) for analyzing changes or trends in technology or service.

Of course, there are many techniques for information visualization. However, GTM is a probabilistic approach to mapping multidimensional data space onto a low-dimensional latent space and vice versa (Bishop et al., 1998), the contributions to identify new opportunities in technology or service are twofold. Firstly, GTM provides a grid-based two-dimensional map. Accordingly, the blank grid is easily detected as a vacuum, requiring no special subjective judgment. Secondly, GTM promotes objective and automatic interpretation by using the inverse mapping function. Since GTM is capable of inverse mapping i.e., mapping the low-dimensional-latent space into the original data space.

GTM-TT, which is an extension version of GTM, is a mapping and visualization technique for multivariate time series data that allows for a dynamic analysis of data. This technique can be applied to the analysis of trends for the following reasons. First, GTM-TT can contribute in terms of

providing time-based clustering, which can provide evidence of time-based technology or service trends. Second, GTM-TT can contribute to the identification of the dynamic change path of technology or service. That is, GTM-TT provides the change path that can describe the time-based pattern after time-based clustering. This information can provide clues for analyzing the technology or service evolution from analyzing the interrelated patterns within service information.

These two methodologies, the GTM and GTM-TT, have been proved to be viable and beneficial in implementing information visualization. They are also considered promising and effective for identification of new opportunities and analysis of trends in technology or service dealt with in this dissertation.

1.2 Purposes

The overall purpose of this dissertation is to develop and apply GTM-based maps for area of technology and service using the information visualization techniques including GTM and GTM-TT. The dissertation is composed of three studies, each of which is concerned with each of three subjects. The objectives of the three studies are as follows.

In this research, discovering patent vacuums to identify new technology opportunities using GTM-based patent vacuum map is conducted.

A patent map has been widely used for identifying the possibilities and opportunities for new technology (Grandstrand, 1999). Since patents are

useful sources of knowledge about technical progress and innovative activity (Basberg, 1987; Grilliches, 1990; Jaffe et al., 2000; Ernst, 2003; Li et al., 2009), the patent map is a guaranteed useful proxy measure for technological power (Park et al., 2005), has been employed as the representative tool used to grasp diverse features of individual patents and identify complex relationships among patents (Yoon et al., 2002). Since patent maps are presented in visual forms such as charts, tables, or graphs, significant amounts of technological information can be acquired in informative and easy ways. More importantly, patent maps have been employed to identify patent vacuums, which are regarded as an unexplored area of technology that deserves intensive investigation for new technology development. In previous studies, two representative types of patent maps have been used for identifying patent vacuums: a principle component analysis (PCA)-based patent map (Lee et al., 2009) and a self-organizing map (SOM)-based patent map (Yoon et al., 2002).

However, two significant limitations exist in both types of patent maps. The first limitation originates from detecting patent vacuums from the patent map. In previous studies, patent vacuums in the patent map have been detected by the subjective ways, depending on the knowledge and experience of researchers. Since there is no clear standard for detecting vacuums, they have been characterized as the relatively sparse or empty areas in the patent map. Thus, no alternative exists, and patents vacuums must be identified in this work by the subjective judgment of researchers. Consequently, patent vacuums might be detected differently depending on each researcher's knowledge and experience, even in a single patent map.

The second limitation constricting previous patent maps corresponds to the interpretation of identified vacuums. After detecting patent vacuums, the vacuum should be interpreted as a real-world technological opportunity, which is as a key part of patent vacuum mapping. However, the interpretation has relied on manual work by researchers, such as investigating the surrounding patents of target vacuums. Therefore, the interpretation of patent vacuums possesses an inevitable weakness in regards to efficiency and effectiveness. In terms of efficiency, an ample amount of time and effort must be devoted to interpreting the patent vacuum as real-world technology. In terms of effectiveness, quite naturally, a significant subjectivity problem arises since the interpretations vary depending on the knowledge and experience of researchers.

Accordingly, this study proposes a GTM-based patent vacuum map which aims to automatically detect and interpret technology vacuums. Since GTM is a probabilistic approach to mapping multidimensional data space onto a low-dimensional latent space and vice versa, the contributions to the detection and interpretation of patent vacuums are twofold. Firstly, in regards to detection, the GTM-based patent vacuum map provides a grid-based two-dimensional map in which each patent is mapped into the relevant grid. Accordingly, the blank grid is easily detected as a vacuum, requiring no special subjective judgment. This means that GTM can overcome the problem of subjective detection observed in traditional patent mapping, providing objective methods for the detection of patent vacuums in a patent map. Secondly, in regards to interpretation, the GTM-based patent vacuum map

promotes objective and automatic interpretation by using the inverse mapping function. Since GTM is capable of inverse mapping i.e., mapping the low-dimensional-latent space into the original data space, the identified patent vacuums are automatically and objectively transformed to the original dataset. Thus, GTM can cope with the manual and subjective interpretation of patent vacuums, providing automatic and objective interpretation. Therefore, using GTM as a means to identifying patent vacuums overcomes two problems observed in traditional patent vacuum maps: subjective detection and interpretation of patent vacuums.

Unlike the above first study deals with technology area, second study concerned with identifying new service opportunities through derived service vacuums using GTM-based service vacuum map.

Recently, services have dominated most developed economies: a great deal more than half of these countries' gross domestic product is in the service sector (Pilat, 2000). During the last two decades, competition between service companies has been harsh due to the deregulation and globalization of markets, the internationalization of service firms, rapid technological evolution, more mature expectations of customers, and so on (Stevens and Dimitriadis, 2005; Jaw et al., 2010). Predominantly after the rise of the smartphone, consumers have been relying on mobile phones for much more than voice communication (Benbunan-Fich and Benbunan, 2007). According to a report by LogicaCMG, about one-fifth of all mobile phone users worldwide have downloaded into their handsets contents such as directions, weather reports, stock prices, and other types of information (LogicaCMG,

2005). This information is called mobile application services, which cover technological information as well as market information. These application services have greatly increased quantitatively as well as qualitatively. In an economic environment where service largely determines the success of a firm, the timely design and development of new services with creative and innovative ideas are essential for a company's survival (Shen et al., 2000).

Unlike the sector of manufacturing, the service sector encompasses the common features of intangibility and simultaneous consumption; especially, service sectors typically do not apply rigorous process design standards prior to the introduction of new services (Fisk et al., 1993). Therefore, in order to maintain a competitive advantage in the service market, a company should actively seek creative ways to generate new and differentiated services that satisfy customers' expectations (Zhang et al., 2005). In this respect, a service company's main challenge is how to explore opportunities to create new services, which are the precursor of commercial success, as time and effort are invested in the earliest stages of service innovation. This research defines new service opportunities as services that have not been developed or launched yet and but have many chances to be developed. Despite the strategic and technological importance of new service opportunities, little academic research has focused on a quantitative and systematic approach to discovering new service opportunities. In response, this research addresses three questions related to the lack of quantitative and systematic research for discovering opportunities in terms of new service development (NSD).

Therefore, this study proposes an approach for developing and utilizing GTM-based service vacuum maps for automatically and objectively discovering service vacuums. This study tries to analyze the massive information data regarding service intelligence and identifies the service vacuums by using information visualization. Its data source is Apple's AppStore, one of the most important and innovative data sources for new mobile services. This work can contribute to the research themes for service engineering in terms of employing and utilizing the service information data to identify service vacuums.

Lastly, identifying trends of service using GTM-TT service trend map is dealt with.

Recently, the business focus has been increasingly shifting from manufacturing-oriented to service-oriented systems (Yang, 2007; Toivonen and Tuominen, 2009). This change has already taken place in many industries, and the trend is expected to continue as future economic growth is foreseen to come predominantly through services-based businesses (Pilat, 2000; Oliva and Kallenberg, 2003; Gebauer and Friedli, 2005; Toivonen and Tuominen, 2009). This phenomenon is especially true of mobile services where a lot of services are simultaneously generated and many are short-lived. The rise of the smartphone has led customers to use many more services compared with traditional phone use, and the competition in this service industry has grown much tougher (Benbunan-Fich and Benbunan, 2007; Lin, 2011). With the rapid technological advancements, a lot of new services will be generated and consumed, and this phenomenon is expected to continue.

In the wake of this quantitative and qualitative expansion of services, the need to analyze trends of services has been also increased. Due to the explosive increase of services, service companies face difficulties analyzing the patterns and trends in services. However, analysis of dynamic patterns can provide service companies with significant implications regarding the identification of service trends and clues for the development of viable new services. Therefore, the identification of service trends cannot be neglected.

The use of a visualization tool, more specifically, a map, fits this purpose. The use of maps has been receiving great attention as a way to visualize massive amounts of information (Chen, 2003; Ware, 2004; Keller and Tergan, 2005). The use of maps can result in unexpected discoveries, deepened understanding, and new ways of thinking, eureka-like experiences, and other intellectual breakthroughs (Chen, 2010). The use of maps has a strong advantage in that it can allow for the organization of massive service data into a clear and simple visual form that can provide insights into trends and patterns of service evolution, which implies that the role of the map fits for analyzing the trends of services.

However, despite the significance of the map, research into the development of dynamic service maps has been immaterial. While some research has been suggested regarding service maps (Kwak et al., 2010; Kim and Park, 2010; Song et al., 2010), most of this research concerns using maps as a form of static analysis. Kwak et al. (2010) suggested using a service map as a method for transforming non-spatial data to graphical representations for exploring new service opportunities by unifying the service features and

customer needs through principal component analysis (PCA). With a similar context, a user-centric service map was also proposed (Kim and Park, 2010) to identify new service opportunities by matching the context between a potential-needs dictionary and existing services. However, these studies still rely on investigating current status of service, providing a static map. In terms of dynamic map, Song et al. (2010) suggested a dynamic service map that considers service evolution based on product utilization. However, the mapping method in this study still remains qualitative and subjective. Therefore, based on the existing research, we propose that it is important to address the question of how to develop a dynamic service map to identify the service trends in an automatic and objective way.

So, this study proposes an approach for developing and using GTM-TT service trend maps consisting of a service cluster map and a service sequence map for identifying the service trends automatically with an objective perspective. Of course, though it is very hard to identify and understand whole service trends, service clusters on the service cluster map and the sequence of service change in a period on the service sequence map can contribute to identify service trends. GTM-TT is a mapping and visualization technique for multivariate time series data that allows for a dynamic analysis of service data. This technique can be applied to the analysis of service trends for the following reasons. First, GTM-TT can contribute in terms of providing time-based clustering, which can provide evidence of time-based service trends. Second, GTM-TT can contribute to the identification of the dynamic change path of services. That is, GTM-TT provides the change

path that can describe the time-based pattern after time-based clustering. This information can provide clues for analyzing the service evolution from analyzing the interrelated patterns within service information. Therefore, GTM-TT is used to develop a service map for analyzing and monitoring the trend of service. In this study, the Apple AppStore, one of the important and innovative data sources for mobile application services, has been used as the data source for service information.

1.3 Scope and framework

The three studies address the development of GTM-based maps using GTM and GTM-TT, respectively, which are the information visualization techniques and core methodologies in this dissertation. As well, the application of GTM-based maps is determined in each study according to the purpose. The overall scope of the dissertation can be defined as follows.

First of all, implementing GTM-based maps has two main purposes: identifying vacuums, and analyzing trends. Therefore, two types of GTM-based maps are provided in this dissertation: The GTM-based vacuum map, and the GTM-TT-based trend map. With respect to the area of application, the first study is dealt with technology using patent data and the second and third studies are carried out for service using AppStore data. In terms of the dynamics, the type of analysis can be classified into two types: static analysis for the first and second studies, and dynamic analysis for third study.

Above two different types of GTM-based maps are basic building

blocks of this dissertation. Figure 1-2 and Table 1-1 shows the scope of the three studies of this dissertation.

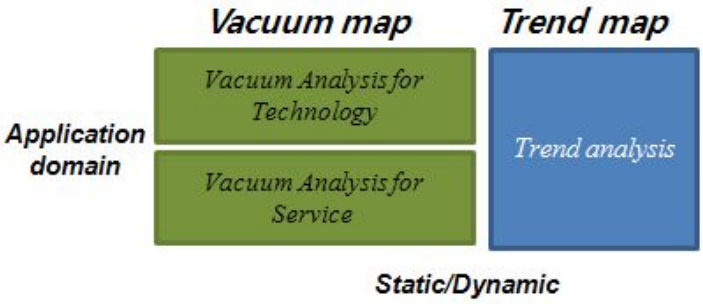


Figure 1-2 Overall scope of dissertation

Table 1-1 Scope of dissertation

Studies	Area	Purpose	Methodology	Data	Analysis
1. Identifying vacuums for technology	Technology	Vacuum	GTM	Patent	Static
2. Identifying vacuums for service	Service	Vacuum	GTM	AppStore	Static
3. Analyzing trends of service	Service	Trend	GTM-TT	AppStore	Dynamic

1.4 Dissertation outline

This dissertation is composed of five chapters and the remainder is organized as follows. Chapter 2 provides background of this dissertation both from theoretical and methodological aspects. Theoretical background provides basic information for understanding the proposed technology map and defining the scope of the dissertation. Explanations of the text mining, GTM and GTM-TT are given as methodological background. Chapter 3 and 4 are main bodies of the dissertation. Basically, the dissertation is organized according to the purpose in implementing information visualization. Three studies are included in both chapters, and each study encompasses its own introduction, proposed approaches to solving problems, and conclusions. Figure 1-3 depicts the overall structure of main body of this dissertation. Finally, the thesis ends with conclusions in Chapter 5. Conclusions include summary, contributions, and limitations, and directions for future research.

Figure 1-3 Overall structure of main bodies of dissertation

Chapter 2. Background

2.1 Theoretical background

2.1.1 Information visualization

Many businesses are using web services to increase information availability and to exchange services among different applications and systems, so research on the potential value of web services has increased (Ray and Ray, 2006). Accordingly, web services themselves can serve as information warehouses. However, their vast quantity of information presents an obstacle. Too information limits interpretation, organization, and/or utilization of information (van der Pijl, 1994; Faia-Correia et al., 1999). Therefore, researchers have proposed information visualization, among other methodologies, for overcoming these limitations.

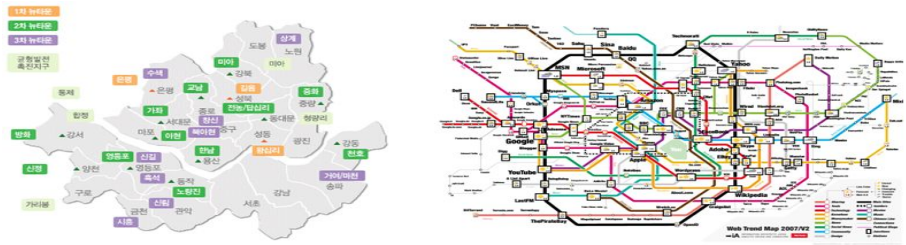
The primary objective of information visualization is to construct a process that describes and explores specific information through graphical representation (McKim, 1980). It is an interdisciplinary research field that integrates an understanding of domain knowledge, data-analytic methods, and computer graphic techniques to enhance cognition. Thus, this field is characterized by representation that is computer-supported, interactive, and visual (Card et al., 1999). When large amounts of data are too unmanageable

to yield critical implications, visualization methods, including maps and networks, help researchers to grasp the overall nature of the data quickly and clearly (Keller and Tergan, 2005; Ware, 2004); these methods also provide new insight into the domain (Chen, 2003). As voluminous data are now being generated in databases, recent studies have placed more emphasis on information visualization for the sake of organization. Visualization can also apply to the service field, as the advanced capability of the smartphone offers many opportunities for developing novel services for both the users and the developers.

The process of information visualization is comprised of seven discrete steps (Yoon, 2010). First, it begins with problem formulation, which clarifies the objective and focus of the visualization. Second, relevant data are collected from appropriate and accessible data sources. Third, the raw data collected during the preceding step are transformed into an appropriate format. Fourth, data mining techniques and multivariate statistical analyses are conducted to derive relationships among the data. Fifth, visualization tools represent the relationships through various visual forms according to the dimension and shape of the data. Data analysis and visualization are strongly related: one provides input information for the other, and vice versa. Sixth, analysts interpret the visualized outputs by investigating graphical forms and their indices. While analysts can derive significant implications from maps or networks by observing the distribution of information, an index presents a quantitative measurement that visual forms cannot offer. Finally, users or experts with domain knowledge or experience must verify the validity and

performance of the visualization.

So far, various types of representation methods, such as trees and networks, have been proposed for different purposes (Shneiderman, 1996). These two representative examples operate as follows. The tree approach is employed to represent hierarchical relationships, whereas the network representation method is applied when a simple tree structure is insufficient for representing complex relationships. The matrix format decomposes a subject into several components by using rows and columns (Yoon, 2010). The charts can help to break down an argument into different perspectives (e.g., information flows and proportions) (Eppler and Burkhard, 2007). Previous studies applied visualization for technology and service areas are described in Table 2-1. Visualization systems based on GTM and GTM-TT belong to a two-dimensional map. Such methods display several maps derived over a large collection of documents, allowing analysts to grasp clearly the overall picture as well as the associations between individual data, understand easily without expertise on statistics, and identify vacuums and trends, intuitively. Figure 2-1 shows the examples about utilities of map for identifying vacuums and trends.



(a) Map of real estate development (b) Map of web trend

Figure 2-1 Examples about utilities of map

Table 2-1 Visualization for technology

Type	**Purpose and usage**	**Authors**
Tree	Identifying characteristics, structuring and understanding entire landscape	Zhang, 2008; Yoon and Kim, 2011
Map	*Grasping big picture and associations between individuals*	*Phaal et al., 2005, Lee et al., 2009; Yoon et al., 2002*
Network	Showing direct relationships	Karki, 1997; Choi et al., 2009
Matrix	Decomposing a subject into sub-components	Yoon and Park, 2007; Yang et al., 2011
Curve	Analyzing life cycle, forecasting, and technology level	Martino, 1983; Nieto et al., 1999

Table 2-2 Visualization for service

Type	**Purpose and usage**	**Authors**
Ontology	Overview of service concept, characteristics	Bianchini et al., 2006; Yang and Chung, 2006; Sensoy and yolum, 2007
Blueprint	Representing a sequential view of service process	Hara et al., 2009; Kingman-Brundage, 1994; Holdford and Kennedy 1999; Shostack, 1982
Matrix	Investing quality of service and developing new service	Arbnor and Bjerke, 1997; David, 1998
Chain	Identifying service actors and focusing on roles, grouping and relations	Morelli, 2002; Bao et al., 2009
Map	*Clustering and providing big picture*	*Del-Hoyo et al., 2009*

2.1.2 Trend analysis

2.1.2.1 Concept of trend analysis

Trend analysis has been employed to analyze the current trend of business, technology, or customers. The purpose of trend analysis can be summarized as follows: to identify the overall pattern of change in an indicator over time, to compare one time period to another time period, to compare one geographic area to another, to compare one population to another, and finally, to make future projections (Rosenberg, 1997). Generally, mathematical or statistical techniques are used to gather time series data to identify future trends. The basic assumption of trend analysis lies in the fact that past conditions and trends will continue in the future more or less unchanged (Wu et al., 2011).

The strength of trend analysis lies in its data-based approach, which offers substantial and data-based forecasts of quantifiable parameters. However, this strength might be also a shortcoming since a significant amount of good data is required to provide a reasonable result for trend analysis. Moreover, trend analysis works only for quantifiable parameters, and is vulnerable to cataclysms and discontinuities (Wu et al., 2011).

To prepare a trend analysis, a series of conceptual issues must be addressed before analyzing and interpreting trend data. These issues include the sample size, which is the number of time periods being examined, the presence of extreme observations or outliers, or the availability of numerator

and denominator data (Rosenberg, 1997).

2.1.2.2 Methods, tools, and techniques for trend analysis

Techniques for trend analysis vary in sophistication, from simple to complex techniques (Wu et al., 2011). The basic methodology for trend analysis is the bibliometric analysis on the basis of literature metadata and information. Bibliometric analysis has been widely used in many fields, such as pharmacology and pharmacy journals, global stem cell research, research on mental health in the workplace, and nanotechnology innovation systems (Wu et al., 2011).

Extending from the bibliometric analysis, more advanced techniques have been suggested, such as visualization methods. Visualization methods are preferred data mining methods because they often offer superior results compared to other conventional techniques (Westphal and Blaxton, 1998; Kim et al., 2008). Visualization methods haves great power in terms of information representation since they provide means for quick and easy knowledge retrieval. For example, high level managers who make technology investment decisions find visualization methods more useful than conventional methods, such as text or tables (Ganapathy et al., 2004; Kim et al., 2008).

Among the many visualization techniques, the most widely used method is the map. Maps offer a visual representation of an area, highlighting the relationships between related elements, such as objects, regions, and

themes. In most cases, maps are represented in either a two-dimensional space or three-dimensional space. Maps can be applied to provide a quick and easy way to capture important, information. Even though the concept of a map starts from a geographical context; it is widely used in the management disciplines in terms of identifying management trends.

To develop a map, text-mining is frequently used to discover previously unknown information by automatically extracting information from various unstructured data (Delen and Crossland, 2008; Wu et al., 2011). In particular, mapping is used to identify the technology trajectories in the scientific areas, where text-mining has been vigorously employed (Liang and Tan, 2007), and technology-intelligence tools also have assisted experts to make strategic technology plans., Moreover, many tools for technology intelligence have been suggested to identify technology trends and provide technological insights through content analysis of technical documents (Yoon and Kim, 2011).

2.1.2.3 Application of trend analysis

Trend analysis is conducted for many different fields. In the scientific fields, a trend analysis model has been used that forecasts the trends of innovation, based on the different stages of innovation and the various data sources associated with each stage of innovation (Martino, 2003; Wu et al., 2011). For basic research, academic papers with science citation indexes can be used as data sources; whereas, for applied research, those with an engineering index

can be used. During the development stage of an innovation, US patents can be a good data source for conducting technological trend analysis. Other forms of media can be used in a similar way, such as newspaper abstracts, which might provide useful implications for trend analysis for various applications, and articles from the business and popular press, which could be good sources of information for measuring the social impact of ideas and innovations.

While the knowledge carriers required in technological development change with each stage, the two most important knowledge carriers for trend analysis with bibliometric and patent methods are scientific papers and patents (Wu et al., 2011). To identify knowledge hidden in the text, text-mining has been employed as a popular method to uncover information from data sources.

2.2 Methodological background

2.2.1 Text mining

The main objective of text mining is to discover previously unknown knowledge from a large collection of texts (Losiewicz et al., 2000). It employs various methods in the research fields of information retrieval, data mining, and statistics to extract valuable information from a large collection of documents (Feldman et al., 1998). The distinct strengths of text mining lie in its ability to transform unstructured documents into structured forms and handle voluminous data. As a massive amount of unstructured documents is now being published on the web, recent studies have placed more emphasis on the use of text mining. It has been employed in various research areas, such as web engineering (Yang and Lee, 2004), technology management (Lee et al., 2011; Yoon and Park, 2007), service management (Kim et al., 2008), and expert systems (Weng and Liu, 2004).

The basic procedure of text mining is composed of four steps. The first step is data collection and preprocessing. Secondly, the structural elements at the levels of words, phrases, and sentences are identified through a linguistic analysis of domain-related and situation-related elements (Moens, 2006). To this end, a variety of structures could be employed for different purposes. The structural elements extracted at this step should be rearranged to consider abbreviations, synonyms, and singular/plural forms, since several different words may represent the same meaning and some extremely

common words are of little value to text mining (Manning et al., 2008). Thirdly, the sampled texts are mapped as syntactic components within this template. To this end, many algorithms, such as the ones for parsing, pattern recognition, and syntactic analysis, could be incorporated to improve the performance of the process. Finally, the results are evaluated by experts who have domain knowledge or experience.

2.2.2 Generative topographic mapping (GTM)

2.2.2.1 Basic concept of the GTM

The GTM is potentially one of the most useful techniques for patent mapping, compensating for the shortcomings of the aforementioned techniques. The GTM was first suggested by Bishop, Svensén, and Willams (1998), proving to be a creditable alternative to the SOM in terms of using a probabilistic method based on Bayesian theory. This method has been utilized across a range of practical applications such as classification, clustering, and visualization (Hogo, 2010). Andrade, Nasuto, Kyberd, and Sweeney-Reed (2005) applied the GTM to the clustering and visualization of motor unit action potentials. Yang and Zhang (2001) proposed the approach to customer data mining and visualization for grouping customer needs using the GTM.

The GTM overcomes most of the limitations found in both the PCA and SOM. Because GTM can present data on each grid, a blank grid is automatically detected as a vacuum. In contrast, the PCA has difficulty in

automatically detecting vacuums due to ineffective visualization. GTM effectively overcomes the limitations of the SOM, including the lack of theoretical proof and over-fitting, by a probabilistic method based on Bayesian theory. This technique also allows a nonlinear relationship between the latent and observed variables (Andrade et al., 2005). In short, the GTM provides a nonlinear mapping algorithm based on probabilistic theory. A major characteristic of GTM is an "inverse mapping" algorithm based on Bayes' theorem, which transforms data in the latent space (as a posterior event) into elements in the data space (as a prior event). In regards to patent vacuums, inverse mapping enables the automatic and objective interpretation of patent vacuums because keywords of core technology in patent vacuums can be extracted. The comparisons of the GTM in patent map with the PCA and the SOM are summarized in Table 2-3.

Table 2-3 Comparisons of the GTM with the PCA and SOM in patent map

	PCA	SOM	GTM
Mathematical backbone	Linear algebra (eigenvalue, eigenvector)	Artificial intelligence (learning process)	Statistics (Bayes' theorem)
Advantage	Meaningful dimensions Theoretical evidence	Forms of discrete nodes	Automatic identification of patent vacuums Forms of discrete grids
Disadvantage	Ineffective visualization Subjective identification of patent vacuums	Absence of general proofs of convergence Ambiguous dimensions Subjective identification of patent vacuums	Ambiguous dimensions

2.2.2.2 The algorithm of the GTM

The underlying principle of the GTM is simple: latent variables are transformed into the data space based on a probability distribution which is estimated in terms of a mean (x), a weight matrix (W), and a noise (β) as shown in Figure 2-2. The x indicates a reduced data vector in the latent space; RL and t represent an observed data vector in the data space, R^D.

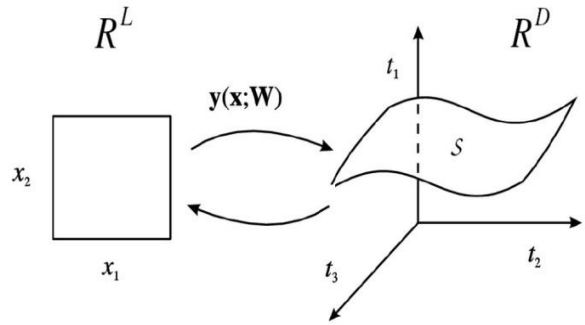


Figure 2-2 Basic concept of GTM (Bishop et al, 1996)

A Gaussian mixture distribution is used as the probability distribution in order to identify closeness in terms of distance between transformed latent data y(x) and observed data t as described in Eq. (1). If data t is close to y(x) in the data set, the probability of $p(t|x)$ becomes higher.

$$p(t|x,W,\beta) = N(y(x,W),\beta) = \left(\frac{\beta}{2\pi}\right)^{-D/2} \exp\left\{-\frac{\beta}{2}\sum_{d}^{D}(t_d - y_d(x,W))^2\right\} \quad \cdots\cdots \quad (1)$$

y: transformation function, x: latent variables,

t: data variables, D: dimension of t,

y(x, W): transformed x into data set, β: noise

The distribution of data in the **t**-space in Eq. (2) is expressed by an integration over the **x**-distribution according to law of total probability.

$$p\left(t\middle|W,\beta\right)=\int p\left(t\middle|x,W,\beta\right)p(x)dx \quad \ldots\ldots\ldots\ldots \quad (2)$$

However, it is difficult to deduce $p(t)$ because it is continuous distribution. To overcome this issue, the delta function is applied to discrete $p\left(t\middle|x\right)$. Applying the delta function also adopts the SOM concept in which the so-called the GTM grid is fabricated and allows the opportunity to locate data on the discrete nodes of predetermined regular GTM grids. Figure 2-3 is an example of a 3 by 3 grid in the latent space and the data space. The final probability distribution with the delta function is transformed by Eq. (3).

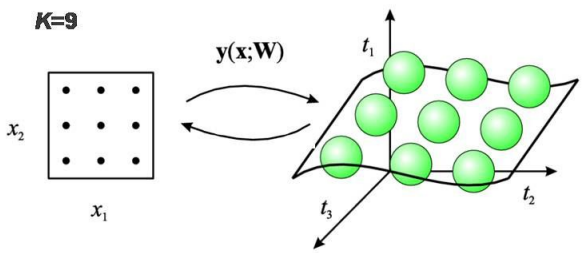


Figure 2-3 Mapping regular grids into the data space

$$p\left(t\middle|W,\beta\right)=\frac{1}{K}\sum_{k}^{K}p\left(t\middle|x_K,W,\beta\right) \quad \ldots\ldots\ldots\ldots \quad (3)$$

K: the number of grid pointers, x_K : a grid point in the latent space

The parameters, which are the weight matrix and noise, are estimated by the Expectation-Maximization (EM) algorithm. After fitting the GTM to a data set, the observed data points can be assigned to latent variables through estimating the probability of a data point, which is generated by a latent point using Bayes' theorem in Eq. (4).

$$p\left(x_k \middle| t_n, W, \beta\right) = \frac{p\left(t_n \middle| x_k, W, \beta\right) p(x_k)}{\sum_{k'} p\left(t_n \middle| x_{k'}, W, \beta\right) p(x_{k'})} \quad \ldots\ldots\ldots\ldots \quad (4)$$

Finally, the observed data can be moved to the latent space, and vice versa as in Eq. (5).

$$y(x, W) = \Phi(x) W \quad \ldots\ldots\ldots\ldots \quad (5)$$

$\Phi(x)$: M fixed basis functions of latent variables, W: D x M matrix

For the details, see the Appendix A.

2.2.3 Generative topographic mapping through time (GTM-TT)

Manifold learning models attempt to describe multivariate data in terms of low-dimensional representations, often with the goal of allowing the intuitive visualization of high-dimensional data (Olier and Vellido, 2008). GTM has been introduced by Bishop, Svensén, and Williams (1996) as one of the statistical machine learning models that provides a principled alternative to the self-organizing map (SOM) algorithm of Kohonen (1982). Unlike the

SOM, the GTM model defines a genuine probability density and thereby overcomes many of the limitations of the SOM (Bishop et al., 1997). Its probabilistic setting has enabled various subjects, such as missing data imputation (Carreira-Perpiñan, 2000; Vellido, 2006), discrete data modeling (Bishop et al., 1998; Girolami, 2002), and robust outlier detection and handling (Bullen et al., 2003; Vellido, 2006).

The GTM-TT is one such extension of GTM for the exploratory analysis of multivariate time series (Bishop et al., 1997) by performing simultaneous time series clustering and visualization. Multivariate time series are not independent identically distributed (i.i.d.) data (i.e., multivariate data corresponding to nearby times will be highly correlated). Therefore, the standard definition of the GTM can only provide a rough approximation to its proper modeling. To deal with this limitation, Hidden Markov Model (HMM) is used in the GTM-TT. As a result, GTM-TT can be understood as a topology-constrained HMM. Parameter estimation can be accomplished in GTM-TT by maximum likelihood using EM algorithm, in a similar fashion to HMMs: details can be seen in a paper by Olier and Vellido (2008).

The GTM-TT is appropriate for analyzing trends with multivariate time series, where the data are correlated. Further, the two-dimensional map provided by GTM-TT is very useful for analyzing and understanding complex trends. For the details, see the Appendix B.

Chapter 3. Identifying vacuums: The GTM-based vacuum map

3.1 The GTM-based patent vacuum map for identifying technology vacuums

3.1.1 Overall research framework

Figure 3-1 depicts the overall research framework, which consists of several stages. Firstly, patent documents related to technology under consideration are collected from the U.S. Patent and Trade Office (USPTO) database. Secondly, text mining tools and experts extract keywords from the documents. Since patents are composed in natural language forms, the documents must be transformed into structured data; in other words, documents must be transformed into arrays of keyword vectors in order to be interpreted, a process regarded as data preprocessing. Thirdly, the patent map is developed by employing GTM. In this GTM-based map, patent vacuums are identified as the blank areas in the map. Lastly, the identified patent vacuums are again transformed to the original keyword vectors using the inverse mapping function of GTM in order to interpret the meaning of patent vacuums.

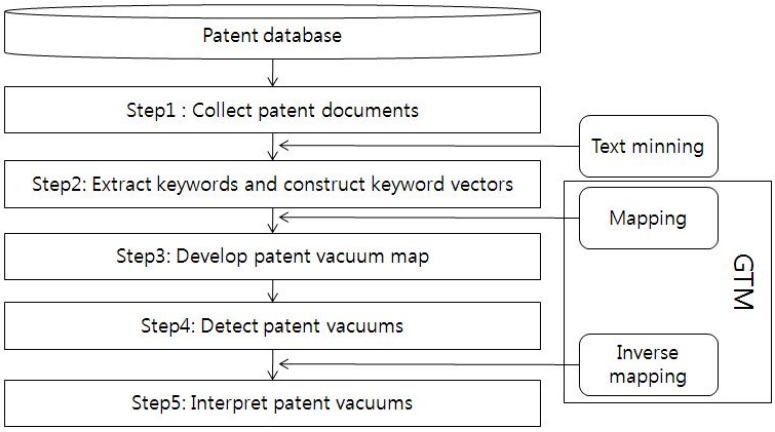


Figure 3-1 Overall research framework

3.1.2 Detailed processes

3.1.2.1 Data preprocessing

The United States Patent and Trademark Office (USPTO) database serves as the data source for collecting patent documents. Thus, patents of interest are searched on the USPTO and collected by Java software that was developed for collecting the patent documents. Keywords are extracted from the collected patent documents in order to construct the keyword vectors, which are used for patent mapping. Since keywords can be considered as brief summaries of a text (Ercan and Cicekli, 2007), in the context of this study, it is possible to think of them as a summarized description of technologies. In this research, Text analysis 2.32, which is a text mining tool, is used for keyword extraction. If we use all extracted keywords from the text mining tool to construct the keyword vectors, information loss can be reduced;

however, the explanatory power decreases due to the complexity of the keyword vectors. Thus, only the most significant keywords should be selected. During this process, the keywords that have no explanatory power, such as device, user, and system, are excluded. The keyword vector is then constructed using Java software, as shown in Table 3-1. The column represents the keywords extracted from the previous step, and the row represents each patent. The value of matrix is either the frequency of keyword occurrence, or the binary value representing the existence of a keyword for each of the patents. See details in Section 3.2.2.2.

However, since the objective of this study is to identify the patent vacuums and not to investigate patent trends, binary keyword vectors were employed instead of the frequency of keyword occurrence vector. For instance, if patent 1 has keyword 1 and keyword n, two fields are filled with "1", respectively.

Table 3-1 The format of keyword vector

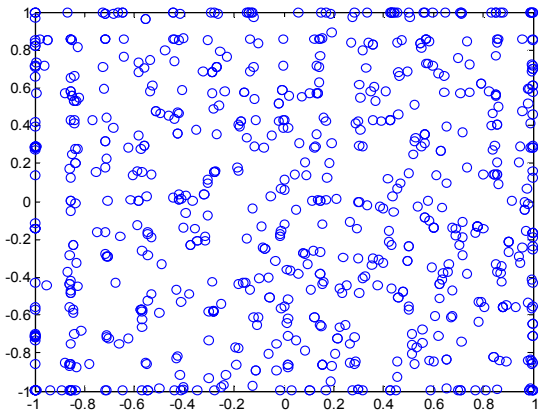
	Keyword 1	*Keyword 2*	...	*Keyword n-1*	*Keyword n*
Patent 1	1	0	...	0	1
Patent 2	1	1	...	1	1
Patent 3	0	1	...	0	0
...	1	0	...	0	0
Patent m	0	1	...	0	0

3.1.2.2 Development of GTM-based patent vacuum map

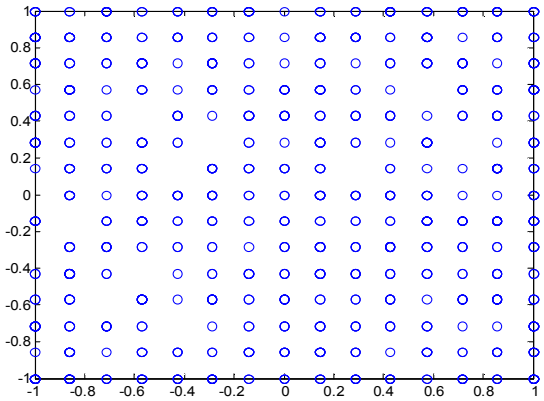
Subsequently, the GTM-based patent vacuum map is developed from the constructed keyword vectors. If fifty keywords are extracted, each keyword vector has fifty dimensions. This presents difficulties in both visualizations and interpretations. Therefore, it is necessary to visualize the vectors in two-dimensional space in order to indentify the patent vacuums using GTM algorithm.

Model parameters must be defined prior to the employment of GTM. The parameters consist of, but are not limited to, the number of latent points and basis functions, the width parameter of the basis functions, the weight regularization factor, and the number of iterations. Svensén (1998) explained that parameters must be chosen individually for each problem. The basis function parameters, which control the smoothness of the mapping, are typically chosen to be radially symmetric Gaussians in which the centers are distributed on a uniform grid in latent space. The width parameter of the basis functions determines the distance between the basis functions. In addition, it is also necessary to select latent space sample points. Note that if there are few sample points in relation to the number of basis functions, the Gaussian mixture centers in the data become relatively independent, and the desired smoothness properties may be lost. Having a large number of sample points, however, increases computational cost. And there is one parameter to set for training: the weight regularization factor. This parameter governs the degree of weight decay applied during training. In practice, because a finite number

of latent and data points are used, a small degree of weight regularization is generally advisable as this prevents the weights from growing very large. Otherwise, smoothness imposed by the basis function parameter could result. Accordingly, the GTM-based patent vacuum map is constructed, as illustrated in Figure 3-2. Figure 3-2(a) shows the posterior-mean projection of the data in the latent space and Figure 3-2(b) shows the posterior-mode projection of the data. In particular, the posterior-mean projection does not precisely identify patent vacuums, but it indicates the original location of the patent and the distance between patents. On the other hand, since all data points are mapped at each latent grid in the posterior-mode projection, the patent vacuums are discovered more clearly than the posterior-mean projection. Each 'o' in Figure 3-2(b) represents a keyword vector mapped at one of the latent points in the posterior-mode projection, and the blank latent points clearly indicate the patent vacuums. Therefore, the posterior-mode projection is more suitable for identifying patent vacuums.



(a) The posterior-mean projection of the data



(a) The posterior-mode projection of the data

Figure 3-2 An example of the GTM-based patent vacuum map

3.1.2.3 Detection of patent vacuums

In the GTM-based patent vacuum map, patent vacuums are identified as the blank areas in the map. As shown in Figure 3-3, the blank grid, which is represented by an X in red, are identified as patent vacuums. Since the GTM-based patent vacuum map mainly consists of grids and each patent is located at each grid, the blank grid is intuitively identified as a vacuum. Thus, manual work conducted by researchers is unnecessary for identifying patent vacuums in the GTM-based patent vacuum map.

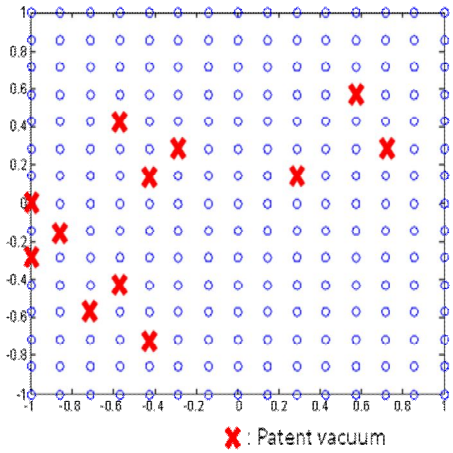


Figure 3-3 An example of patent vacuums

3.1.2.4 Interpretation of patent vacuums

Patent vacuums are then transformed to the original keyword vector through inverse mapping following identification in order to provide the original meaning of the patent vacuum. The characteristics of inverse mapping, which differentiate GTM from other latent variable models, enable projection from the latent space into the data space (Bishop et al., 1998). Thus, manual and subjective interpretation of identified patent vacuums limiting previous attempts is eliminated by the automatic and objective interpretation of identified patent vacuums through the inverse mapping function of GTM. Consequently, keyword vectors are identified by inversely mapping (Eq. (5)) patent vacuums in latent space into new vectors in data space as illustrated in Figure 3-4. The keyword vector fields of patent vacuums are transformed as binary values by threshold value, so the value '1' implies that the value is

over threshold value determined by analyst as illustrated in Figure 3-4. Since there is no definitive method in determining the threshold value, it is determined depending on the purpose of research. That is, if threshold value is low, identified patent vacuums comprise of many keywords.

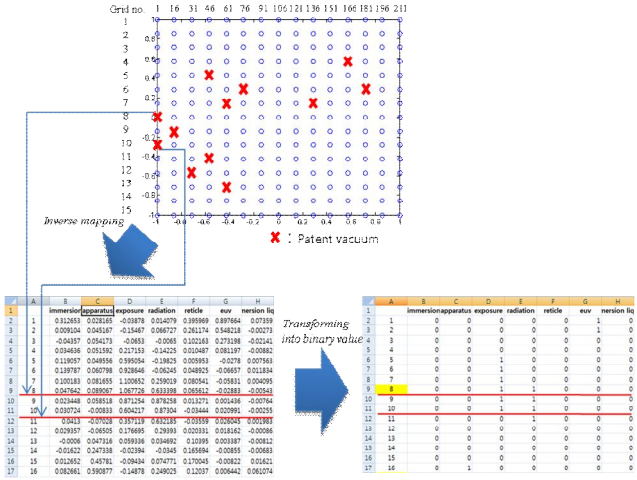


Figure 3-4 An example of inverse mapping

3.1.3 Case study: lithography technology

In this section, a case study of lithography technology demonstrates the applicability of the proposed approach. Lithography technology is regarded as one of the most critical aspects in the semiconductor manufacturing processes (Stulen and Sweeney, 1999; Harriott, 2001). Lithography technology-related patents were selected for two main reasons. Firstly, a large amount of new lithography technology has been examined in order to survive in the highly competitive semiconductor manufacturing environment. Consequently, the

demand for new lithography technology has been increasing continuously (Fay, 2002). While lithography process control is becoming increasingly complex, and lithography technology progresses toward smaller feature sizes, specifications are tightening, demanding better lithography process control (Janakiram and Goernitz, 2005). Secondly, the number of collected patents related to lithography technology is suitable to mine underlying information and develop patent maps. Therefore, lithography technology is considered an appropriate subject matter for illustrating the proposed approach. For more details about lithography and all the technologies that support this field, the reader is referred to two textbooks (Smith, 1998; Levinson, 2001).

3.1.3.1 Data collection

As mentioned above, lithography technology patents are the underlying source for data presented in this study. The United States Patent and Trademark Office (USPTO) database serves as the data source for collecting patent documents. Patents contain diverse information, such as patent number, title, abstract, registered year, inventor, assignee, citation, claim, and description. To collect the lithography-related patents, patents which have the word 'lithography' in each title, abstract, and claim parts were selected. As a result, 754 lithography-related patents with a reference period between 1976 and 2009 were collected.

3.1.3.2 Data preprocessing

Since most text-mining algorithms use keywords for expressing the context of the document (Yoon and Park, 2004), this study regards keywords as the data source that represent the characteristics of patents. The abstract of each patent provided the venue from which keywords were extracted because the abstract is vital literature explaining important information that the patent author wishes to convey. With the aid of Text Analysis 2.32, keywords were extracted automatically. Afterwards, keywords that have no explanatory power were eliminated according to the expert judgment of officials in the semiconductor field. As a result, a total of 40 keywords were extracted. These keywords are described in Table 3-2.

Table 3-2 Extracted keywords

No.	Keyword	No.	Keyword	No.	Keyword	No.	Keyword
1	Immersion	11	Wavelength	21	Grid	31	Modulation
2	Apparatus	12	Photomask	22	Deflection	32	Refraction
3	Exposure	13	Temperature	23	Maskless lithography	33	Defocus
4	Radiation	14	Modulator	24	Pupil	34	Polarization
5	Reticle	15	Interferometer	25	Detector	35	Contamination
6	Euv	16	Deflector	26	Dose	36	Sigma
7	Immersion liquid	17	Calibration	27	Lens	37	Curvature

8	Laser	18	Immersion medium	28	Fresnel	38	Alignment
9	Axis	19	Pellicle	29	Aberration	39	Bandwidth
10	Path	20	Numerical aperture	30	Frequency	40	Pulse

Subsequently, data mined from each patent was transformed into a keyword vector consisting of binary values. If a specific keyword was included in each patent, the corresponding vector field was assigned a value 1; otherwise, a value 0 was assigned. A Java program was used for constructing the keyword vector. As a result, keyword vector was constructed, as illustrated in Figure 3-5.

Keyword	:	Immersion	Apparatus	Exposure	Radiation	Reticle	•••	Pulse
Patent 1	:	(1	1	1	1	1	•••	0)
Patent 2	:	(1	1	0	1	0	•••	1)
⋮				⋮				
Patent 754	:	(1	1	1	0	0	•••	0)

Figure 3-5 Keyword vector construction

3.1.3.3 Development of GTM-based patent vacuum map

After data preprocessing, GTM was employed to develop the GTM-based patent vacuum map for identifying patent vacuums. Prior to developing the GTM-based patent vacuum map, parameters must first be defined. The main model parameters were defined by sensitive analysis as follows:

A GTM model comprised of a 14-by-14 square grid of latent points in two-dimensional space. The model utilized 81 Gaussian basis functions in which the center of each function was located on a 9-by-9 square grid in the latent space. Both grids were centered about the origin in the latent space. The basis functions had a common width of 1.5 times the shortest distance between two neighboring basis functions. The model was initialized using PCA, and trained for 10 iterations of the training algorithm. The weight regularization factor governs the degree of weight decay applied during training was 0.001. With above parameters, GTM-based patent vacuum map is developed using MATLAB R2008a with GTM toolbox developed by Svensén (1998) as shown in Figure 3-6.

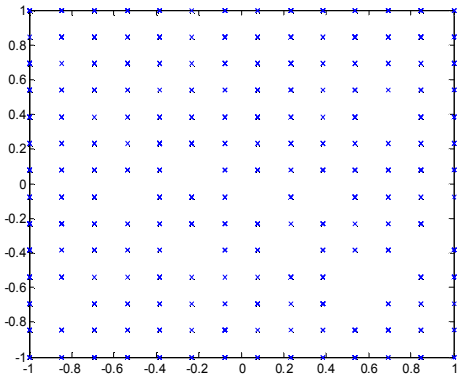


Figure 3-6 GTM-based patent vacuum map (The posterior-mode projection)

3.1.3.4 Detection of patent vacuums

Detecting patent vacuums was conducted through posterior-mode projection

since posterior-mode projection provides a clearer representation of patent vacuums. The blank grid is identified as the patent vacuum. Figure 3-7 shows the patent vacuum identified from the GTM-based patent vacuum map. A total of 13 patent vacuums of 169 latent points were discovered through posterior-mode projection.

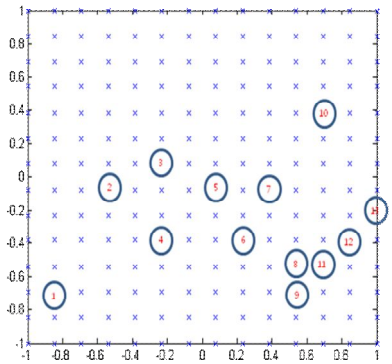
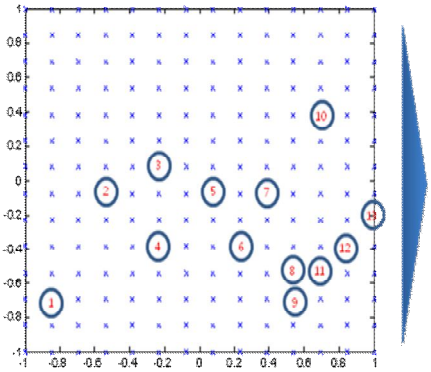


Figure 3-7 Patent vacuums identified from GTM-based patent vacuum map

3.1.3.5 Interpretation of patent vacuums

Inverse mapping was conducted in order to interpret the meaning of the identified patent vacuums. Each vacuum in Figure 3-7 was transformed into the keyword vector as a means to represent the characteristics of each vacuum. The results of the inverse mapping are represented in Figure 3-8.



A	B	C	D	E	F	G	H
	immersion	apparatus	exposure	radiation	reticle	euv	nersion liq
1	0.100927	0.829164	0.709588	0.096982	0.04564	0.023643	0.018761
2	0.061276	0.627586	0.863242	0.128809	0.227482	0.158402	-0.00945
3	-0.0053	0.745934	1.141932	0.548884	0.543886	0.30308	0.020904
4	0.003433	0.5387	0.929759	0.602487	0.927248	0.209397	0.0031
5	0.001518	0.517445	0.95178	0.056888	0.560499	0.111716	-0.01742
6	0.014282	1.077975	0.616263	0.246613	0.496176	-0.06349	-0.00681
7	-0.07892	0.026011	0.932023	0.519861	0.050945	-0.03344	0.000431
8	0.0762	0.887467	0.818623	0.236081	0.571688	0.102355	0.000866
9	0.088849	0.441416	0.75358	0.215359	0.688054	0.130216	-0.00115
10	0.508061	0.837005	0.902489	0.916147	0.550358	0.281488	0.393269
11	-0.02147	0.892802	0.957173	0.052325	0.564359	0.085607	-0.01174
12	-0.08553	0.764497	1.028191	0.269401	0.61448	0.106781	-0.01009
13	0.153047	0.926599	1.04002	0.921665	0.983157	0.445722	0.219095

Figure 3-8 The result of inverse mapping

Since the original keyword vector is composed of binary values, the result of inverse mapping should also be composed of binary values. Therefore, each field of keyword vector obtained from the inverse mapping was filled with a 0 or 1 value according to the threshold, as shown in Table 3-3. The threshold value was set as 0.4 considering the level of variance of the keyword vector.

In the case of the first identified patent vacuum, three keywords existed, “apparatus”, “exposure”, and “lens” since three 1 values are located in the 2nd, 3rd, and 27th of keyword vector field. This means that each patent vacuum identified from GTM-based patent vacuum map can be interpreted in the level of keywords, which is the form of the original dataset.

Table 3-4 shows the final result of the vacuum interpretation, illustrating the result of inverse mapping of each patent vacuum. This result was validated whether or not a patent with these specific keywords existed in the USPTO. The interpreted patent vacuums provide valuable evidence for

locating unexplored areas of technology that deserve intensive investigation for new technology development. However, despite the contribution, it should be clearly understood that this result only shows the patent vacuum identified from a technical perspective, not from the "real-world" patent vacuum. Since this result is obtained by inverse mapping of GTM only, an in-depth level of feasibility test is required in a practical setting.

Table 3-3 Keyword vector of patent vacuums

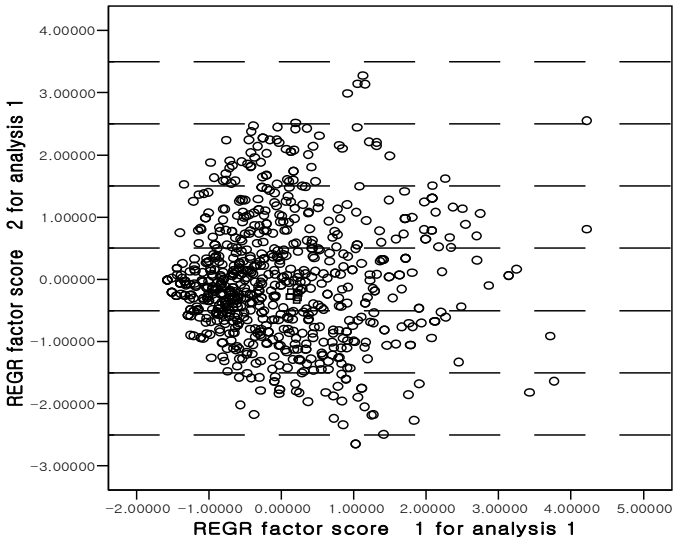
Vacuum No.	***Keywords***																																								*Sum*
	1	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***	***13***	***14***	***15***	***16***	***17***	***18***	***19***	***20***	***21***	***22***	***23***	***24***	***25***	***26***	***27***	***28***	***29***	***30***	***31***	***32***	***33***	***34***	***35***	***36***	***37***	***38***	***39***	***40***	
1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3
2	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	3
3	0	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	6
4	0	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7
5	0	1	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	6
6	0	1	1	0	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	9
7	0	0	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	9
8	0	1	1	0	1	0	0	0	1	1	0	0	1	0	0	0	1	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	12
9	0	1	1	0	1	0	0	0	1	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	1	0	0	0	12
10	1	1	1	1	1	0	0	1	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	1	0	0	0	14
11	0	1	1	0	1	0	0	1	1	1	1	0	1	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	16
12	0	1	1	0	1	0	0	1	1	1	1	0	0	1	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	0	0	0	1	0	1	17
13	0	1	1	1	1	1	0	1	1	1	1	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0	1	0	0	1	0	1	0	1	0	0	0	1	0	0	18

Table 3-4 The final result of vacuum interpretation

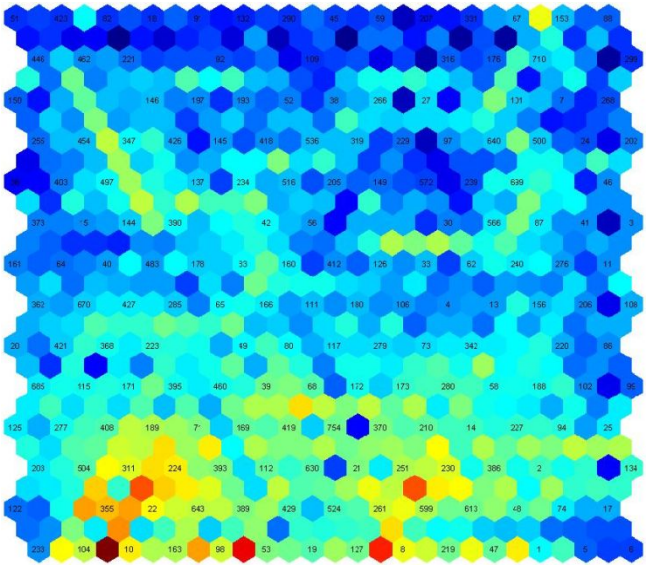
patent vacuum No.	Keywords
1	Apparatus, exposure, lens
2	Apparatus, exposure, alignment
3	Apparatus, exposure, radiation, reticle, wavelength, alignment
4	Apparatus, exposure, radiation, reticle, laser, axis, calibration,
5	Apparatus, exposure, reticle, axis, wavelength, lens
6	Apparatus, exposure, reticle, laser, axis, path, wavelength, lens, alignment
7	Exposure, radiation, laser, axis, path, wavelength, grid, lens, frequency
8	Apparatus, exposure, reticle, axis, path, temperature, calibration, numerical aperture, grid, dose, sigma, alignment
9	Apparatus, exposure, reticle, axis, path, temperature, interferometer, calibration, dose, lens, sigma, alignment
10	Immersion, apparatus, exposure, radiation, reticle, laser, axis, wavelength, temperature, interferometer, lens, aberration, refraction, curvature
11	Apparatus, exposure, reticle, laser, axis, path, wavelength, temperature, modulator, calibration, numerical aperture, maskless lithography, dose, sigma, alignment, pulse
12	Apparatus, exposure, reticle, laser, axis, path, wavelength, modulator, deflection, maskless lithography, dose, lens, frequency, modulation, polarization, alignment, pulse
13	Apparatus, exposure, radiation, reticle, euv, laser, axis, path, wavelength, temperature, interferometer, calibration, detector, lens, frequency, refraction, polarization, alignment

3.1.4 Discussions

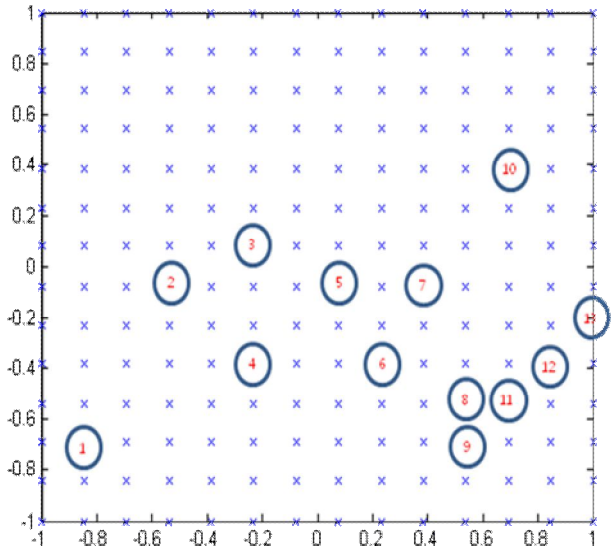
The GTM-based patent vacuum map needs to be validated for practical use in the field, and to show that this can be another alternative comparing to PCA-based patent map and SOM-based patent map. For this, the same patent data set (754 lithography-related patents) was used for PCA-based patent map and SOM-based patent map. Figure 3-9 shows PCA-based patent map, SOM-based patent map, and GTM-based patent vacuum map, respectively.



(a) PCA-based patent map



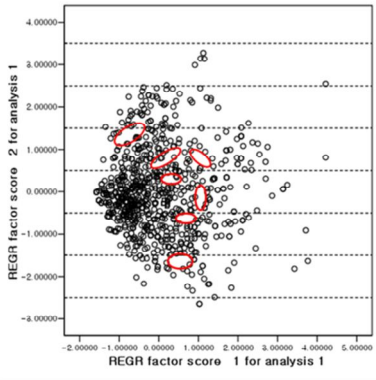
(b) SOM-based patent map



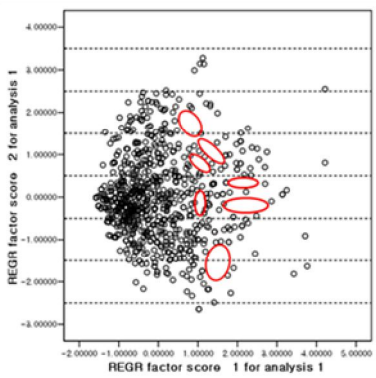
(c) GTM-based patent vacuum map

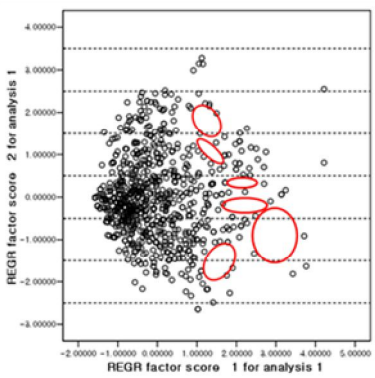
Figure 3-9 Three distinctive patent maps

As mentioned above, PCA-based patent map and SOM-based patent map have two main limitations in terms of detection and interpretation of patent vacuums. Firstly, in terms of detection, patent vacuums might be detected differently depending on each researcher's knowledge and experience in both patent maps, even in a single patent map as shown in Figure 3-10 and 3-11. Sparse areas in the PCA-base patent map in Figure 3-10(a) are considered as patent vacuums. However, patent vacuums represented ellipses might be changed according to researchers' judgments as shown in Figure 3-10 since the definition of sparseness varies depending on the researchers. In Figure 3-9(b), each node is colored depending on the median distance to its neighbors based on a reference vector. Those nodes which belong to a 'dense' region of the map will have a bright color. Thus, the darker the color is, the longer the distance to the neighbors is. Therefore, researchers must judge which area is a vacuum by the color scale and the location of patents in SOM-based patent map. So, it also causes the same limitation with PCA-based patent map in respect to detection of patent vacuums as shown in Figure 3-11. However, patent vacuums in the GTM-based patent vacuum map are automatically detected using grid-based visualization since a blank grid is considered as a patent vacuum as shown in Figure 3-9(c).



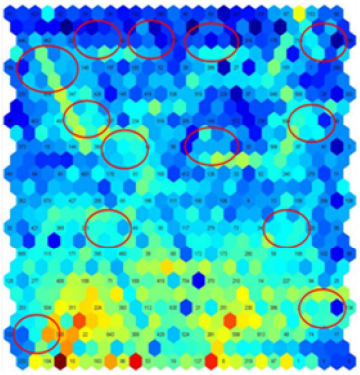
(a) Researcher 1



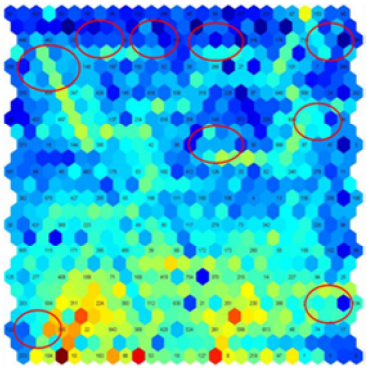


(b) Researcher 2 (c) Researcher 3

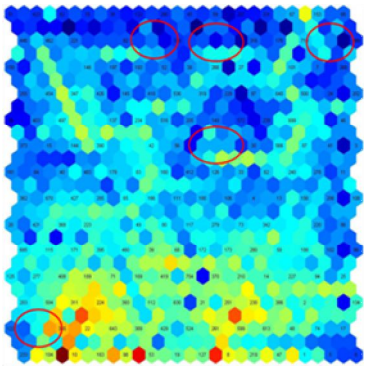
Figure 3-10 Patent vacuums depending on researchers in PCA-based patent map



(a) Researcher 1



(b) Researcher 2



(c) Researcher 3

Figure 3-11 Patent vacuums depending on researchers in SOM-based patent map

Secondly, in terms of interpretation, both patent maps should investigate all the surrounding patents of target patent vacuums for interpretation of patent vacuums since there is no function for interpretation of the meaning of the patent vacuum. Therefore, lots of time and efforts are devoted for interpretation of patent vacuums and interpretations vary depending on the knowledge and experience of researchers. For instance, the

surrounding patents of target patent vacuum expressed arrow in Figure 3-10(a) and 3-11(a) are shown in Table 3-5. It means that 28 and 33 surrounding patents in PCA-based patent map and SOM-based patent map should be manually investigated by researchers for interpretation of a patent vacuum, respectively.

Table 3-5 List of surrounding patents of target patent vacuum

Surrounding patents of target patent vacuum					
PCA-based patent map			*SOM-based patent map*		
4692579	5742065	6674086	3701391	6427703	7067222
4924257	5756234	6724001	4606803	6716563	7081948
4985634	5786601	6817602	4677042	6800428	7091502
4987311	6090528	6968253	4881257	6849856	7129024
5068884	6127272	7096127	4969169	6879380	7189981
5111491	6255038	7295288	4969169	6887630	7283205
5187726	6369398	7579606	5313068	6897076	7304775
5204886	6387572	7631289	5326979	6953644	7332734
5424549	6465796		5426686	6958804	7414701
5719698	6522433		6373071	7026098	7435978
					7438997
					7521689
					7625513
Total number					
		28			33

On the contrary, GTM-based patent vacuum map overcomes this limitation through the function of inverse mapping so that keyword vectors as means of each patent vacuum are automatically identified as shown in Table 3-5. Although identified keyword vectors may not fully explain the technology, those provide enough clues to systematically explore technological vacuums. Consequently, the GTM-based patent vacuum map is more appropriate for identifying patent vacuums among lots of patents since it can automatically and objectively detect and interpret patent vacuums.

3.2 The GTM-based service vacuum map for identifying service vacuums

3.2.1 Overall research framework

The overall process of the proposed approach consists of four stages: construction of the database, preprocessing, development of the service map, and exploration of service opportunities, as shown in Figure 3-12. In the first step, the researcher collects the service documents from the service area of interest in order to construct the database (DB) for analysis. Since the service documents collected in this step are unstructured data in the sense that they are expressed in text format, it is necessary for the researcher to transform the unstructured data into structured data for analysis. Therefore, in the second step, the researcher conducts the preprocessing through the text-mining technique to (a) convert the collected documents into arrays of keyword vectors based on keywords extracted from the DB for developing the service map and (b) categorize the extracted keywords for facilitating the interpretation of the final outputs. In the third step, the GTM algorithm is employed to develop the service map through MATLAB. In this manner, each service in the DB is mapped on one grid on the two-dimensional service map. In the fourth step, service vacuums are identified as blank grids in the service map, and the researcher then explores new service opportunities by

transforming the identified service vacuums into the original keyword vectors by using the inverse mapping function of GTM. Finally, new service opportunities are derived from the service vacuums through screening, and they are characterized by the number of adjacent services surrounding each service vacuum. Finally, evaluation follows. These steps are introduced sequentially hereafter.

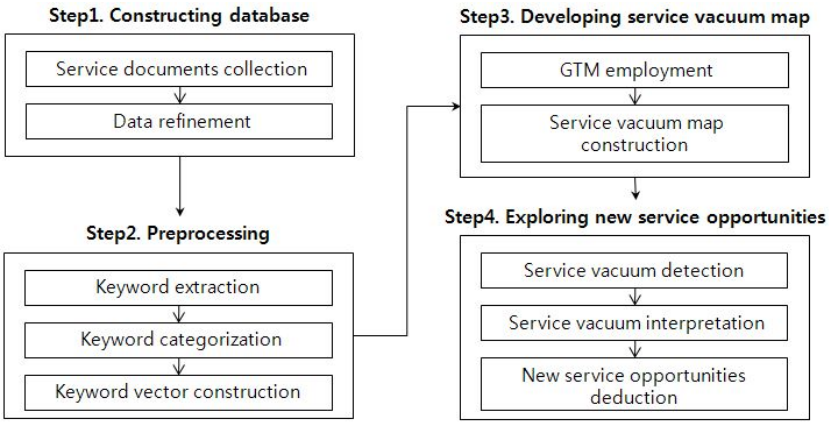


Figure 3-12 Overall research framework.

3.2.2 Detailed processes

3.2.2.1 Step 1: Construction of the database

The first task is to construct the database. For this task, service documents serve as the data that are collected from the service field of interest to be analyzed. Although data related to services can be collected from various data sources, such as reports from service firms, academic papers, books,

newspapers, and survey data, the web has been considered an emerging data source recently, since it offers plentiful and diverse data, and easier computer-based techniques can be applied for collecting and analyzing the data. In fact, data collection plays a key role within analysis, as the data must truly emulate the realities of the system to the levels of accuracy and detail required. Failure to obtain the required data to this level of accuracy and detail implies that the model will not accurately emulate the physical system, thereby producing invalid results and recommendations (Robertson and Perera, 2002). Furthermore, in practice, it is very difficult to collect a lot of data manually. Therefore, this type of research begins by using a computer-based system that runs on JAVA to automatically collect data from the web. Then, experts who have domain knowledge or experience refine the collected data by eliminating unrelated data. However, these documents cannot be analyzed directly, since the service documents collected at this step are unstructured data, that is, they are simply expressed in text format. So, an additional step for transforming them into structured data is necessary.

3.2.2.2 Step 2: Preprocessing

After the data are collected, the keywords of documents are extracted through text mining in order to construct the keyword vectors that are used for developing the service map. Since keywords can be considered as brief summaries of a text (Ercan and Cicekli, 2007), in the context of this study, it is possible to think of them as a summarized description of services. Although

the summary of a text is capable of providing more information about the text than the keywords of a text, a summary may not be suitable for some applications due to the complex structure of sentences. Keywords are therefore not replacements for summarization but alternative summary representations that can be consumed by other applications more easily.

The extraction of keywords involves text-mining software just as much as it does the experts' judgment, since it is difficult to describe all the service characteristics using only a text-mining technique. Consequently, repetitive trials between experts and a computer-based approach are required to define the form and elements of the keyword vector in a service context. The detailed procedure is depicted in Figure 3-13. Text-mining software can yield not only the importance of each keyword but also the relationship between keywords in all the documents through the keyword list and the keyword structure. As a result, compilation of a term-frequency (TF) matrix counting the occurrences of each term in each service document is derived. Term dimensionality is typically reduced by excluding trivial English words, low-frequency terms, and term conflation (such as stemming or lemmatization). The purpose of this reduction is not just computational efficiency but, more important, avoidance of spurious language patterns. Therefore, after executing text mining, the experts arrange the keywords into a hierarchical structure to identify the relationship between keywords. In this way, they can discover overlapping or conflicting keywords and remove them. Then, the keywords are listed in the order of frequency, and the keywords with high frequency are considered in the analysis due to the assumption that

more important keywords appear more frequently. And then, only the keywords with high importance will be selected as the first candidates and undergo the experts' screening process. In the screening process, the role of domain experts is to eliminate meaningless or redundant keywords out of the list of high frequency keywords. In addition, they will include keywords with low frequency but are critical and highly related to the service. To avoid conflicts between experts at the screening process, domain experts repeatedly execute a scoring with 5-likert scale at every decision making process until reaching a consensus. Lastly, these experts verify whether all extracted keywords are appropriate for constructing the keyword vectors. Then, for each service document, the occurrence of the keywords used in the document is assigned to a corresponding vector field so that each service document can be represented by a keyword vector.

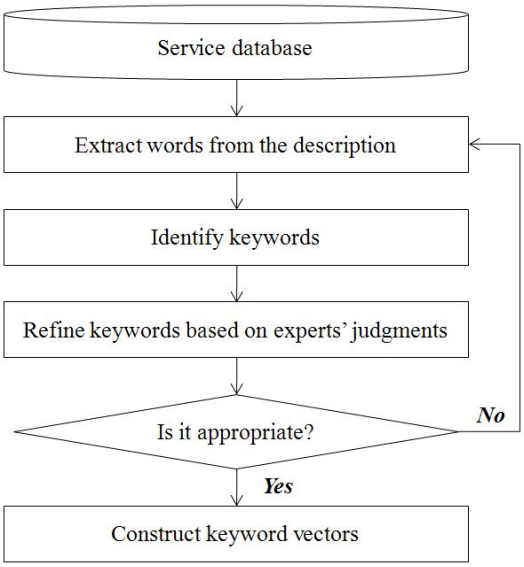


Figure 3-13 Procedure for constructing keyword vectors.

Once the experts choose the keywords, they categorize them into several clusters by matching their features for ease of interpretation of the results of analysis (i.e., service vacuums), since service vacuums are provided as a set of keywords.

A keyword vector is constructed to visualize the relationships among documents that contain service information. The keyword vector helps researchers to discover meaningful patterns in documents (Losiewicz et al., 2000; Yoon et al., 2002), given that the service map is developed on the basis of the keyword vector. In other words, the massive information in collected documents containing information about existing services can be transformed into structured keyword vectors, as shown in Figure 3-14. The present study constructed the keyword vector according to binary values instead of

occurrence frequencies. That is, the binary value for a keyword is assigned to the corresponding vector field on the basis of whether or not the keyword appears, because the main objective of the present study is to not analyze the trends or characteristics of existing services but to identify the service vacuums on the service map. As a result, each service document can be distinguished by a keyword vector.

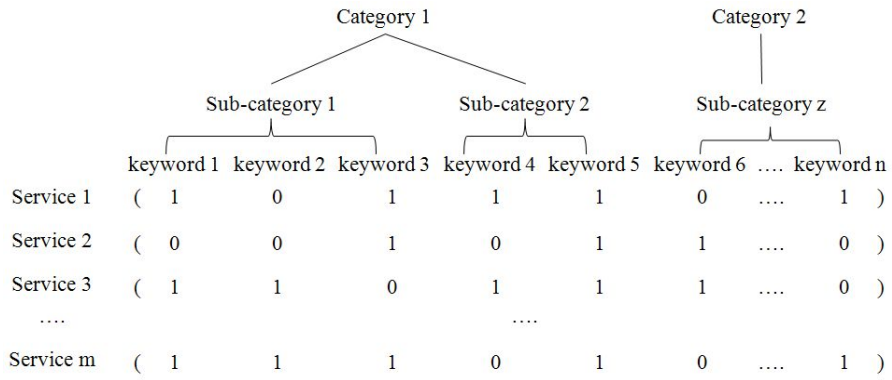


Figure 3-14 Structure of keyword vectors.

3.2.2.3 Step 3: Development of a GTM-based service vacuum map

After data preprocessing, each service is mapped to a rectangular planar surface in order to generate the service map. When various keywords are used to construct a keyword vector, visualization and interpretation are difficult due to the number of dimensions of the keyword vector. Therefore, one of the critical tasks is to reduce the number of dimensions of keywords to an acceptable level. In this research, GTM is employed to develop a GTM-based

service vacuum map with multi-dimensional keyword vectors. By employing GTM, this study was able to develop a service map, which appears in Figure 3-15.

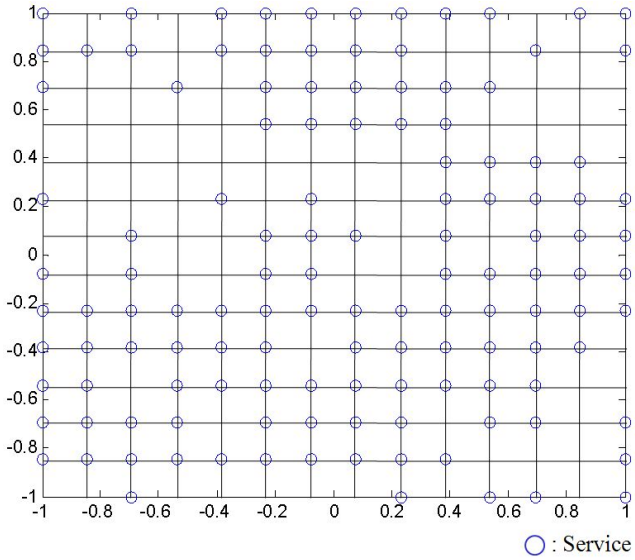


Figure 3-15 Example of the GTM-based service vacuum map.

Unfortunately, it is still difficult to ascertain the meaning of each axis on a GTM-based service vacuum map. However, as the ultimate goal of mapping is to identify service vacuums and not to interpret meanings on the map, there is less need for a description of each axis. When the points on the map are assigned by estimating the probability of a data point (service document) that is generated by the Bayesian theorem, nearby points will have similar features regarding services. This display of data in the service map facilitates an understanding of the structures in the dataset.

3.2.2.4 Step 4: Exploration of new service opportunities

3.2.2.4.1 Identification of service vacuums

In the GTM-based service vacuum map, service vacuums intuitively appear as blank grids on the service map, as shown in Figure 3-16, since the map mainly consists of grids and each service is located at a grid. Thus, manual work on the part of researchers is unnecessary for identifying service vacuums on the GTM-based service vacuum map.

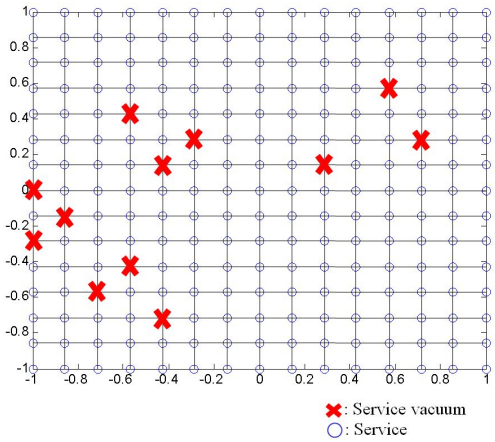


Figure 3-16 Example of identifying service vacuums in the service map.

Once they identify the service vacuums, they can explore the new service opportunities by transforming the identified service vacuums into the original keyword vectors by using the inverse mapping function of GTM. The characteristics of inverse mapping, which differentiate GTM from other latent

variable models, enable researchers to project the latent space onto the data space (Bishop et al., 1998). Thus, the manual and subjective interpretation of the identified service vacuums, which has limited previous attempts, is eliminated by the automatic and objective interpretation of the identified service vacuums through the inverse mapping function of GTM. Consequently, keyword vectors are derived by inversely mapping service vacuums in the latent space onto new vectors in the data space. An example of the procedure for interpreting service vacuums appears in Figure 3-17.

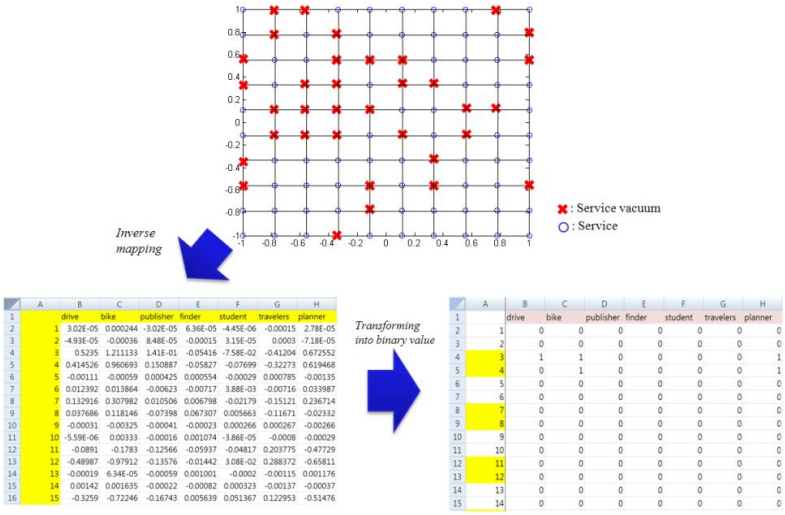


Figure 3-17 Example of interpreting service vacuums.

3.2.2.4.2 Selection of new service opportunities from identified service vacuums

The meaningful and valuable new service opportunities are determined from

the service vacuums through the following analyses: significance analysis and explanation power analysis. The two indexes for significance analysis regarding the importance of keywords and the other two indexes for explanation power analysis are operationally defined, as shown in Table 3-6. Then, based on those indexes, the service vacuums and their service features are analyzed to find new service opportunities.

First, a significance analysis aims at evaluating the significance of service vacuums for future service opportunities. This analysis measures the importance of service vacuums by dividing the keyword frequency of vacuums by the number of keywords in the service vacuum. The number of emerging keywords from the service vacuums is divided by the overall number of emerging keywords. Firstly, the keyword frequency indicates how frequent the particular concept is mentioned in the document. Tiun et al. (2001) established that the higher the frequency, the more important the concept is deemed to be, so the average of a keyword's frequency in a service vacuum is regarded as the indicator for the importance of the service vacuum. This frequency is measured with the importance of service (IoS) index. Secondly, keywords can be categorized into two types: emerging or declining keywords. A service vacuum with a high number of emerging keywords and a small number of declining keywords is expected to be an important service opportunity (Lee et al., 2009). Emerging keywords are the keywords whose appearance frequency in the documents increases over time, and declining keywords appear less frequently as more documents appear in the database. This phenomenon leads to the Newness of Service (NoS) index.

Table 3-6 Indexes to explore new service opportunities.

Purpose	Index	Definition
Identifying new service opportunities regarding service vacuums	EoS (the degree of explanation of the service)	The number of keywords of a service vacuum
	CoS (the degree of completeness of the service)	The number of categories filled with keywords
	IoS(the degree of importance of the service)	The average of keyword frequency in a service vacuum
	NoS(the degree of newness of the service)	The rate of emerging keywords in a service vacuum
Defining the characteristics of new service opportunities	FoS (the degree of feasibility of the service)	The number of adjacent services surrounding a service vacuum

Second, the explanation power analysis evaluates the degree of explanation power of the service vacuums as new service opportunities. The number of keywords and the number of categories filled with keywords, which are respectively measured by the Explanation of Service (EoS) index and the Completeness of Service (CoS) index are utilized for this analysis. The optimal number of keywords and categories to explain a document is still open to debate. However, this analysis assumes that two indexes with high value will have high explanation power of service. That is, in this research, a service vacuum including a few keywords and/or not including any keywords for all categories can be a clue for developing a new service but does not provide enough information to explain a service. Thus, such service vacuums are excluded, and the remaining service vacuums are defined as new service opportunities.

However, even if a vacuum's scores are high on IoS, NoS, EoS, and CoS, it still may not be relevant as an NSO from a managerial or technological point of view. That is, the results may reveal mutually inconsistent or contradictory pairs of keywords representing service components, and such alternative combinations will clearly be managerially or technologically impossible to develop. In the case of cameras, for instance, an analogue camera would not be equipped with a memory card nor would a digital camera require film. Therefore, configurations with "analogue/memory card" or "digital/film" pairs must be removed from the set of alternatives. This part of the process requires domain experts to establish a contradiction matrix that investigates all possible pairs and excludes inconsistent or

impossible configurations from the list of alternatives to be considered.

3.2.2.4.3 Characteristics of new service opportunities

To gasp the characteristics of new service opportunities, experts measure the number of adjacent services for each service vacuum with the Feasibility of Service (FoS) index, as shown in Table 3-6. A service vacuum that is surrounded with a high density of adjacent services is considered a NSO with high feasibility and a competitive edge (Kohonen, 1995; Yoon et al., 2002). Additionally, Kwak et al. (2010) pointed out that a company will have as much chance of gaining competitive advantages through the differential strategy. They also stressed that companies must be cautious of the high risk involved in the creation of breakthrough innovations in service vacuums that are surrounded by low density service areas. In the case that the service vacuum is surrounded by high density service areas, companies will find it easier to develop new service by modifying the adjacent services and this investment presents relatively low risk.

3.2.3. Case study: navigation mobile application service

3.2.3.1 Data collection

Data for navigation application services were collected from Apple's

AppStore homepage[1] which provides information on mobile application services. AppStore is an application portal on which developers upload and distribute their mobile application services with some useful information to analyze. Accordingly, AppStore can serve as the data source for collecting service documents. The information on each mobile application service is derived from the data sources provided by the mobile service providers, such as the category, last changed date, version, price, size, service provider, detailed descriptions, and the types of data sources provided by users, such as reviews. Figure 3-18 provides a list of these sources. Both data source types are useful to the purposes of the present study, because they describe services. That is, this study can use several data or one datum out of a specific data source type or several data from both data source types in the analysis. Since the purpose of this study is to identify new service opportunities through analyzing developed or launched services in terms of service contents and characteristics, it used the detailed description of the data source provided by service providers.

[1] http://www.iphoneappsplus.com/

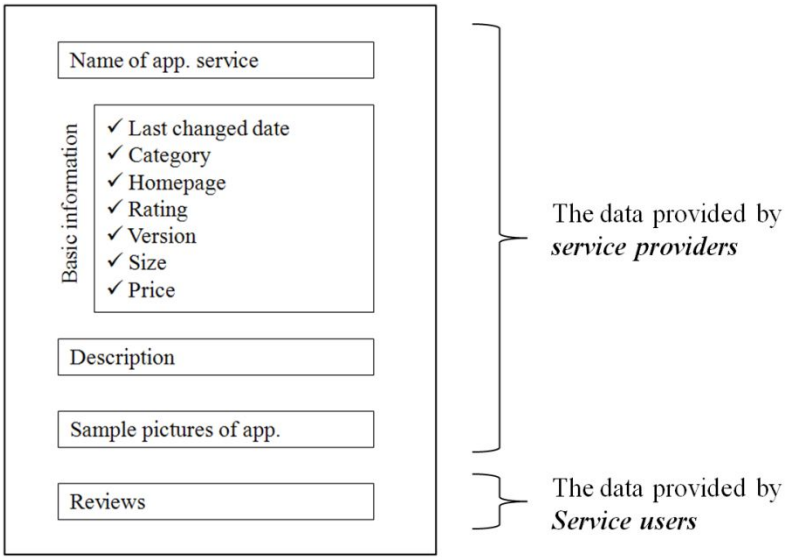


Figure 3-18 Structure of a service document.

We searched for the service documents of interest on the AppStore and collected them via the JAVA software that we developed for collecting the service documents. We employed services related to navigation applications to demonstrate the applicability of the proposed approach. The category of navigation is one of 20 application categories, which also include books, business, education, etc. A total of 2609 navigation application services prior to November 2009 were collected. The reason that we selected navigation application services is two-fold. Firstly, discovering new service opportunities in the sector of navigation application services is more important than other discoveries, since competition among developers is fiercer due to the relatively expensive prices compared to other services and the market has already matured. Secondly, the number of collected documents

related to navigation application services is suitable for mining the underlying information and developing a service map. Hence, navigation application services are considered an appropriate subject matter for illustrating the proposed approach.

3.2.3.2 Data preprocessing

3.2.3.2.1 Extracting keywords

From the set of collected documents, keywords were firstly extracted automatically with the aid of Text Analysis 2.32, which is a text mining tool developed by Megaputer Intelligence Inc. This process is shown in Figure 3-19.

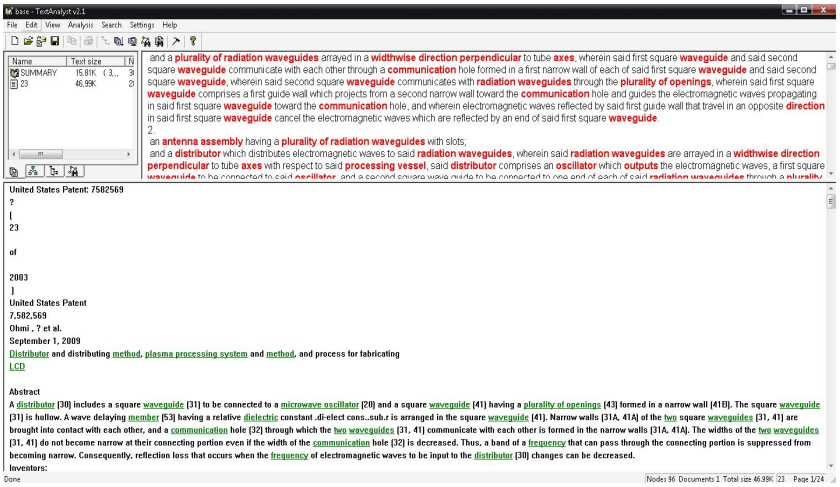


Figure 3-19 Example of keyword extraction using TextAnalyst 2.32.

TextAnalyst processes text automatically as neural networks but produces semantic structures as the end product. TextAnalyst views the original text as a sequence of symbols, taking snapshots of groups of symbols at a time. A hierarchical neural network is derived with the most important attribute being frequency of occurrence. The relationship among symbols also affects the hierarchy. TextAnalyst ignores common words, such as the articles "a," "and," and "the." Through this process, the most important words, phrases, and sentences in text can be extracted.

As a result, a total of 4840 keywords were derived using TextAnalyst. After the elimination of supplementary words such as "system," "device," "function," and so on, expert screening was executed to refine the keywords precisely. Consequently, a total of 272 remaining keywords were used in the analysis as keywords that could thoroughly explain navigation application services. See Appendix C for keyword list.

3.2.3.2.2 Structuring of the extracted keywords

After extracting keywords, we structured them, including their categorization for interpreting the results of analysis, and constructed the keyword vectors for employing GTM by analyzing the keywords of the service documents.

All keywords are grouped into several clusters according to their characteristics by keyword structure that resulted from the text mining and the results of the literature analysis. After categorization, validation was executed by domain experts. Even if more than one data source is used, the only change

is the number and the type of categories, since the categorization is performed based on the level of keywords. As mentioned above, this task helps experts to interpret service vacuums that consist of particular keywords. In this research, a two-level service hierarchy is suggested for the categorization of keywords. At the first level, keywords are categorized according to the dimensions of the user, technology, and information, and each dimension has subcategories at the second level, which are depicted in Figure 3-20.

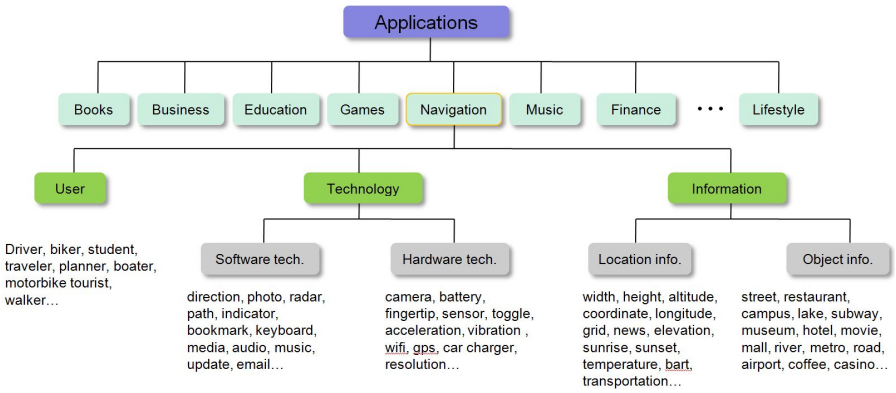


Figure 3-20 Service hierarchy for navigation mobile application services.

The keyword vector is then constructed to be used as input data for constructing a service map. In the keyword vector, the column represents the keywords extracted from the previous step, and the row represents each mobile application service. The keyword vector fields register binary values; thus, '1' for the 'tourist' field in application service 1 means that the keyword, 'tourist,' exists in the service document for mobile application service 1.

3.2.3.3 Developing a GTM-based service vacuum map

A service map is developed by means of the GTM algorithm with the constructed keyword vectors. A multi-dimensional keyword vector is mapped on a two-dimensional planar surface. To develop the GTM-based service vacuum map, this study used MATLAB R2008a with the GTM toolbox provided by Svensén (1988). In practice, a service map is constructed differently according to the parameters used for employing GTM. In other words, the shape of the service map affects the number and features of service vacuums, and the shape changes with different parameters, including the number of grids, the basis function, the number of iterations, and so on. For more details about parameters, we refer the reader to Svensén's (1988) thesis.

Among all types of service maps, an appropriate service map, as shown in Figure 3-21, is selected through a sensitivity analysis of the map's ability to discover new service opportunities. The chosen service map has a proper number of service vacuums for analysis, and each service vacuum is distinguished relatively well from others.

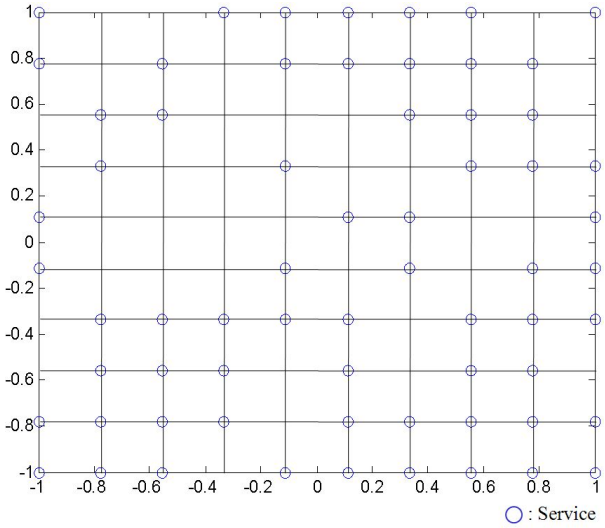


Figure 3-21 A service vacuum map (10-by-10) selected by sensitivity analysis.

3.2.3.4 Exploring new service opportunities

3.2.3.4.1 Detection of service vacuums

On the GTM-based service vacuum map, service vacuums are identified as blank areas on the map. As shown in Figure 3-22, blank grids, which are represented by an X in red, have been defined as service vacuums. Since the GTM-based service vacuum map mainly consists of grids and each service is located at a grid, a blank grid is intuitively identified as a vacuum. Thus, researchers can automate the identification of service vacuums.

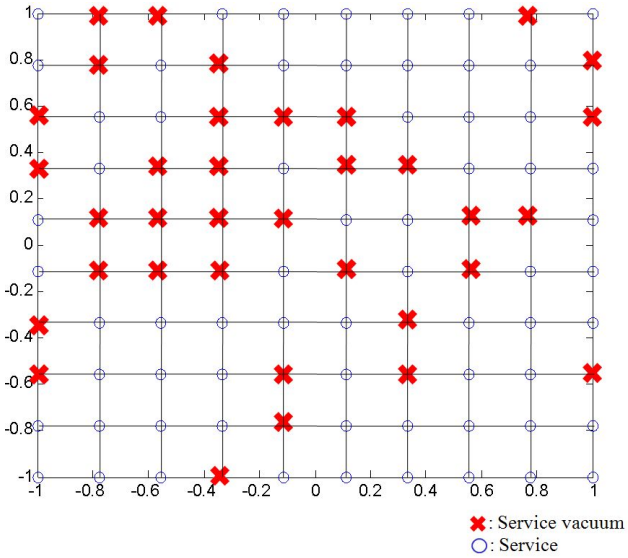


Figure 3-22 Detection of service vacuums in the GTM-based service vacuum map.

3.2.3.4.2 Interpretation of service vacuums

Each service vacuum is then interpreted to discover new service opportunities among the detected service vacuums. Note that one of the advantages of GTM is to provide an inverse mapping function, which can trace what the original data is. Hence, the detected service vacuums can be expressed in the form of keyword vectors as means to represent the description of each service vacuum through the inverse mapping function.

Table 3-7 shows the list of 35 service vacuums, including the number of keywords belonging to each category, keyword frequency, and the number of emerging keywords such as GPS, Wi-Fi, email, bookmark, etc. for each service vacuum. This means that each service vacuum can be an

alternative new service opportunity and its description is expressed by the keywords of that service vacuum. Each service vacuum has been numbered from top to bottom and then from left to right on the service map. To illustrate, service vacuum 33 is a service that includes the keywords “biker,” “publisher,” and “tourist” for the category of “user”; “Wi-Fi” for the category of “hardware technology”; “offline,” “drag,” “guidance,” “tracker,” “prediction,” and “translation” for the category of “software technology”; “city” and “office” for the category of “object information”; and “depth” and “sunrise” for the category of “location information.”

Table 3-7 The list of service vacuums.

Vacuum No.	No. of keywords					*Total*	Keyword frequency	No. of emerging keywords
	User	Technology		Information				
		Hardware	Software	Object	Location			
1	3	3	16	12	4	*38*	*13,177*	*24*
2	2	3	11	9	2	*27*	*11,634*	*18*
3	0	2	4	8	0	*14*	*10,886*	*10*
4	0	2	2	6	0	*10*	*8,174*	*8*
5	0	2	3	3	0	*8*	*10,115*	*6*
6	0	3	3	4	0	*10*	*10,496*	*8*
7	0	3	2	7	0	*12*	*11,128*	*8*
8	0	3	3	8	0	*14*	*9,176*	*7*
9	0	2	3	1	0	*6*	*9,407*	*4*
10	1	2	10	7	2	*22*	*10,807*	*16*
11	1	2	6	8	4	*21*	*9,355*	*14*
12	0	2	4	7	0	*13*	*7,966*	*8*
13	1	3	7	3	1	*15*	*10,160*	*14*
14	3	2	12	3	1	*21*	*9,364*	*18*
15	4	3	9	6	2	*24*	*9,616*	*19*
16	0	3	5	6	3	*17*	*8,197*	*14*
17	0	2	3	4	0	*9*	*7,113*	*8*
18	3	6	9	5	5	*28*	*8,382*	*11*
19	2	2	7	4	1	*16*	*8,379*	*13*
20	0	2	1	2	0	*5*	*9,009*	*4*
21	0	2	2	2	0	*6*	*8,006*	*5*
22	0	2	1	2	1	*6*	*8,990*	*5*
23	0	2	1	2	0	*5*	*5,678*	*3*
24	1	3	3	4	4	*15*	*7,501*	*4*
25	0	1	2	3	0	*6*	*6,453*	*3*
26	0	2	3	2	1	*8*	*2,605*	*3*
27	0	1	1	0	0	*2*	*3,581*	*1*
28	1	4	5	1	0	*11*	*4,405*	*4*
29	0	0	3	1	0	*4*	*4,758*	*2*
30	0	1	1	0	0	*2*	*3,626*	*2*
31	4	7	12	5	5	*33*	*8,039*	*11*
32	0	0	0	1	0	*1*	*2,515*	*1*
33	3	1	6	2	2	*14*	*3,568*	*4*
34	1	2	3	2	2	*10*	*3,445*	*1*
35	1	0	3	8	1	*13*	*2,351*	*4*
Total	*31*	*80*	*166*	*148*	*41*	*466*		

3.2.3.4.3 Identifying new service opportunities

After the detection and interpretation of service vacuums, the next step is to screen the meaningful service vacuums as new service opportunities, since not all service vacuums may be feasible new service opportunities. That is, some service vacuums may appear to be fertile but are truly sterile in terms of their potential value. Screening attempts to find the meaningful service vacuums and these vacuums are determined by their keywords. New service opportunities are determined by the values of the above indexes (EoS, CoS, IoS, and NoS), as shown in Table 3-8. If a service vacuum has higher values for the four indexes, it can be considered a new service opportunity. As a result, 9 out of 35 service vacuums, which are shown in bold in Table 3-8, were identified as new service opportunities, and the final result of the screening is shown in Table 3-9, which displays the list of keywords for each service vacuum. In the case of service vacuum 11, for example, a new service opportunity for drivers is identified, which uses hardware technologies including GPS, Internet connection and software technologies such as email, current location detection, favorite, radar, and video. This type of software provides object information, including street, restaurant, city, country, etc. and location information, including coordinate, latitude, longitude, and overview. This explanation of a new service opportunity using keywords is easy to understand, and it would contribute considerably by providing a strong clue for NSD.

Table 3-8 The result of significance analysis and explanation power analysis.

Vacuum No.	EoS	CoS	IoS	NoS
1	**0.14**	**1.00**	**0.049**	**0.52**
2	**0.10**	**1.00**	**0.043**	**0.39**
3	0.05	0.60	0.041	0.22
4	0.04	0.60	0.030	0.17
5	0.03	0.60	0.038	0.13
6	0.04	0.60	0.039	0.17
7	0.04	0.60	0.042	0.17
8	0.05	0.60	0.034	0.15
9	0.02	0.60	0.035	0.09
10	**0.08**	**1.00**	**0.040**	**0.35**
11	**0.08**	**1.00**	**0.035**	**0.30**
12	0.05	0.60	0.030	0.17
13	0.06	0.80	0.038	0.30
14	**0.08**	**1.00**	**0.035**	**0.39**
15	**0.09**	**1.00**	**0.036**	**0.41**
16	0.06	0.80	0.031	0.30
17	0.03	0.60	0.027	0.17
18	**0.10**	**1.00**	**0.031**	**0.24**
19	**0.06**	**1.00**	**0.031**	**0.28**
20	0.02	0.60	0.034	0.09
21	0.02	0.60	0.030	0.11
22	0.02	0.80	0.034	0.11
23	0.02	0.60	0.021	0.07
24	0.06	1.00	0.028	0.09
25	0.02	0.60	0.024	0.07
26	0.03	0.80	0.010	0.07
27	0.01	0.40	0.013	0.02
28	0.04	0.80	0.016	0.09
29	0.01	0.40	0.018	0.04
30	0.01	0.40	0.014	0.04
31	**0.12**	**1.00**	**0.030**	**0.24**
32	0.00	0.20	0.009	0.02
33	0.05	1.00	0.013	0.09
34	0.04	1.00	0.013	0.02
35	0.05	0.80	0.009	0.09

Table 3-9 The final result of analysis.

NSO No.	Vacuum No.	List of keywords					No. of adjacent services
		User	Technology for		Information on		
			Hardware	Software	Object	Location	
1	1	Driver, biker, planner	GPS, internet connection, camera,	Offline, email, current location, favorite, drag, interactive, calculation, bookmark, statistics, upload, wikipedia, topographic, checkpoint, speech, recorder, playback	Location, area, restaurant, road, lake, movie, company, rental car, mountain, valley, county, ATMs	Altitude, transportation, postcode, public transportation	*7*
2	2	Biker, planner	GPS, internet connection, camera	Offline, current location, favorite, drag, interactive, calculation, bookmark, upload, wikipedia, topographic, playback	Location, area, road, lake, rental car, mountain, valley, county, ATMs	Postcode, public transportation	*83*
3	10	Driver	GPS, internet connection	Offline, email, current location, detection, favorite, radar, interactive, statistics, upload, recorder	Street, restaurant, city, country, road, rental car, harbor	Altitude, vessel	*4*
4	11	Driver	GPS, internet connection	Email, current location, detection, favorite, radar, video	Street, restaurant, city, country, building, food, rental car, drink	Coordinate, latitude, longitude, overview	*1*
5	14	Driver, biker, traveler	GPS, internet connection	Offline, email, current location, online, language, interactive, bookmark, statistics, upload, wikipedia, topographic, recorder	City, road, ATMs	Altitude	*64*
6	15	Driver, biker,	GPS, internet	Offline, current location,	Street, restaurant, road,	Altitude, vessel	*61*

		traveler, tourist	connection, fingertip	favorite, language, interactive, statistics, upload, translation, recorder	region, rental car, harbor		
7	18	Biker, navy navigator, motorbiker	GPS, internet connection, wifi, battery, vibration, car charger	Direction, text, photo, community, feedback, indicator, management, plotter, graphic indicator	Region, speedcam, ocean, snorkeling, stairway	Car, latitude, longitude, elevation, depth	*105*
8	19	Biker, traveler	GPS, internet connection	Offline, online, language, tracker, upload, wikipedia, topographic	Area, city, road, ATMs	Altitude	*63*
9	31	Driver, publisher, planner, government	Internet connection, fingertip, vibration, accelerometer, acceleration, windscreen, sunlight	Current location, text, radar, music, indicator, font, interactive, measurement, dashboard, timer, checkpoint, brightness	Location, region, aircraft, speedcam, harbor	Car, height, altitude, situation, vessel	*170*

The remaining service vacuums are treated as new service opportunities in this study. However, it does not mean that they themselves are concrete new services, since they are composed of only several keywords. Nevertheless, their identification provides the intelligent information that experts need to discover new services and to make decisions. Service designers and decision makers can utilize derived new service opportunities, which are a conceptual combination of service components, to identify new services at the fuzzy stage of the NSD process for idea generation.

The characteristic of each new service opportunity is measured by the value of an index (FoS), which is defined as the number of adjacent services surrounding a service vacuum. To illustrate, the number of adjacent services for service vacuum 31 is 170, which means a relatively high number of similar services have already been developed nearby. If a new service that encompasses the keywords of service vacuum 31 is one of the alternatives that will be developed, the service can be created accordingly with high feasibility and competition in terms of NSD. Service vacuums 2, 18, and 31 are characterized as new service opportunities that have high feasibility and competition, as shown in Figure 3-23.

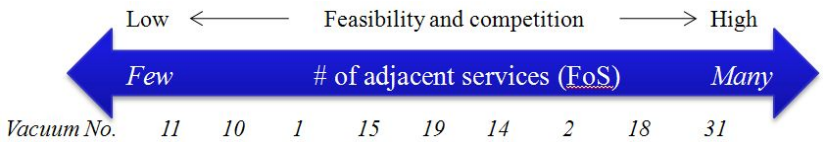


Figure 3-23 Characteristics of the identified new service opportunities.

3.2.3.2 Evaluation of new service opportunities

The applicability of the suggested methodology needs to be evaluated. One way to execute this evaluation is by means of a holdout sample that uses two data sets according to time period: 2008-2009 (the data used to identify new service opportunities) and 2010-2011 (the added data to evaluate recognized new service opportunities), since it is useful to set aside part of the data to assess how well the selected model performs on the holdout data. After the promising new service opportunities are excavated by the first data set, examination as to whether they have been exploited in the second period data set is carried out.

Table 3-10 presents the list of navigation mobile application services in the second data set that fit the derived new service opportunities from the first data set. Four out of nine new service opportunities have been developed and launched since 2010. This shows that valuable but undeveloped mobile application services in one period were realized in the next. The applicability of the suggested approach to the task of forecasting promising new service opportunities can therefore be validated.

Table 3-10 List of realized navigation application services in second period

NSO No.	Mobile application service launched in second period	Description
2	VoxTrek navigation	GPS navigation with text to speech directions and speed camera detection
3	Navigation By TeleNav	Voice guided turn-by-turn directions with spoken street names, auto rerouting, 3D maps with live traffic flow, cheap gas price finder and more
	Team Navigation	GPS monitoring and tracking service for the iPhone
5	iWay GPS Navigation	Voice-guided GPS navigation system with turn-by-turn driving directions and other features
8	Sygic series	Premium, comfortable and safe turn-by-turn GPS navigation app with an intuitive interface, valuable content
	Scout	Make use of topographic maps for hiking adventures or check training data with the integrated speedometer
	Komoot Bike	Outdoor GPS navigation for mountain biking, hiking, mountain climbing and mountaineering

3.2.4 Discussions

The way we used to describe services in this study is constructing keyword vectors using extracted keywords from service data. Therefore, any data of both data source types above shown in Figure 3-18 are also possible to describe services, if it can represent service contents and features. There are several ways of constructing keyword vectors using possible data to describe one service as follows. Firstly, keywords are extracted from the data source provided by service providers, then the data provided by users are used to help select keywords through prioritizing extracted keywords, and vice versa. Secondly, keywords are simultaneously extracted from two kinds of data source types, and then keyword vectors are constructed using extracted keywords. However, it is hard to decide which way is the most appropriate way to describe a service through keywords. It will depend on the type of data, the purpose of study, methodologies, and so on. In the future research, identifying New service opportunities reflecting customer needs considered very important in the service sector and analyzing dynamic evolution pattern of mobile application services will be promising research areas. Accordingly, various data including customer input data such as reviews, rate or dynamic data such as last changed date, version can be utilized.

Because of the emergence of telecommunication, data network, Internet, and mobile Internet, most recently, services are becoming even more virtualized. These virtual services, which are provided via the Internet, are referred to as electronic services (Bouwman et al., 2008). Van de Kar

(2004) defines an electronic service as "an activity or series of activities of intangible nature that take place in interaction through an Internet channel between customers and service employees or system of the service provider, which are provided as solutions to customer problems, add value and create customer satisfaction." The electronic service has characteristics that are found both goods and services, and it also has some unique characteristics of its own (Hofacker et al., 2007), as shown in Table 3-11.

Table 3-11 Distinguishing between goods, electronic services, and services

Goods	Electronic services	Services
Tangible	Intangible, but need tangible media	Intangible
Can be inventoried	Can be inventoried	Cannot be inventoried
Separable consumption	Separable consumption	Inseparable consumption
Can be patented	Can be copyrighted, patented	Cannot be patented
Homogeneous	Homogeneous	Heterogeneous
Easy to price	Hard to price	Hard to price
Cannot be copied	Can be copied	Cannot be copied
Cannot be shared	Can be shared	Cannot be shared
Use equals consumption	Use does not equal consumption	Use equal consumption
Based on atoms	Based on bits	Based on atoms

The major difference between electronic services and many traditional services is the role people play in the service delivery process. An electronic service is not delivered by human but by software programs via computer hardware and communication networks. This has major implications for service characteristics. Electronic services can be accessed

anytime and anywhere. Mobile services, especially mobile application services, are a specific subset of electronic services and defined as services that are offered via mobile and wireless networks. Therefore, in this study, mobile application services are used in terms of this context.

The proposed approach can be used not only in service field, but also in product or technology field. However, the reason why this study focuses on service field is that the quantitative and systematic study for discovering new opportunities in service area is needed because of service data explosion.

Chapter 4. Identifying trends: GTM-TT-based trend map

4.1 The GTM-TT service trend map for identifying trends of service

4.1.1 Overall process

The overall process of the proposed approach mainly includes the following steps: 1) construction of the database, 2) data preprocessing, 3) development of the GTM-TT service trend map, and 4) interpretation, as shown in Figure 4-1. In the first step, the service documents are collected from the service area of interest to construct the database (DB). Since the service documents collected in this step are unstructured data in that they are expressed in text format, it is necessary to transform the unstructured data into structured data for analysis. Therefore, in the second step, preprocessing is conducted by experts through the text-mining technique in which, using text-mining tools, keywords are extracted from the documents, and, subsequently, multivariate time series data from the service documents are generated. For visualization of text data, each document must be transformed into a multivariate time

series data that is made up of quantitative values such as weighted keyword frequency or binary value by the occurrence of keywords and time information of each document. In the third step, the GTM-TT algorithm is employed to develop the GTM-TT service trend map using MATLAB. In the fourth step, the service trends are derived through the interpretation of the GTM-TT service trend map. These steps are introduced sequentially hereafter.

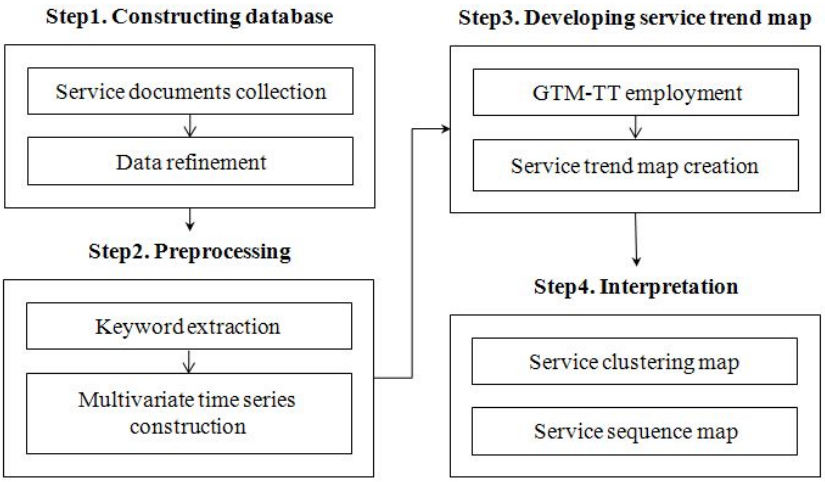


Figure 4-1 Overall process

4.1.2 Detailed procedures

4.1.2.1 Construction of the database

As an initial step, the construction of the database is conducted as follows. First, the search and the selection procedures for the target service area to be analyzed are executed. These procedures are important because they are not only very closely related to the purpose of research, but also affect the overall analysis process. Next, the available data source is determined. Although data related to services can be collected from various data sources, such as reports from service firms, academic papers, books, newspapers, and survey data, the Web has been considered an emerging data source recently because it is a source of plentiful and diverse data and computer-based techniques can be applied for collecting and analyzing the data (Pérez et al., 2008; Gilson et al., 2008). Hence, this study uses Web data related to target services. Finally, service documents in a specific service area are collected. In practice, it is difficult to collect a lot of data manually. Therefore, in this research, a computer-based system using JAVA is used to automatically collect data from the Web. Then, experts who have domain knowledge or experience refine the collected data by eliminating unrelated data.

4.1.2.2 Data preprocessing

Since service documents are expressed in natural language format, it is necessary to transform them into structured forms appropriate for analysis. Most text-mining algorithms use keywords as features for expressing the document (Sebastiani, 2002). Therefore, in this study, the process of keyword extraction is applied to identify keywords, and the extracted

keywords are used to represent services. The detailed steps of the data preprocessing are follows.

In this step, the first task is to extract keywords from collected service documents. Text-mining serves as a tool for analyzing the relationship among service documents. Respective service documents, specifically, are featured by keywords that represent service characteristics. The extraction of keywords relies first, on text-mining technique, and second, on domain experts. After a keyword list is derived from the text-mining analysis, it is necessary to screen the keywords on the list in order to eliminate the ones that are meaningless or irrelevant to the strategic purpose of the analysis. For the purpose, experts' domain knowledge is used, based on which a final keyword set for the following analysis is determined. Consequently, repetitive trials between experts and a computer-based approach are required to define the form and elements of the keyword vector in a service context. Then, for each service document, the occurrence of each keyword is assigned to a corresponding vector field and, as a result, each service document is distinguished by a keyword vector. Once keyword vectors are constructed, they are sorted by first launching date of each service in chronological order to make multivariate time series data that are used as input data for analysis, as shown in Figure 4-2.

	Time	keyword 1	keyword 2	keyword 3	keyword 4	keyword 5	....	keyword n
Service 1	2008/09/03	(1	0	1	1	1	....	1)
Service 2	2008/10/05	(0	0	1	0	1	....	0)
Service 3	2009/01/10	(1	1	0	1	1	....	0)
....					....			
Service m	2009/10/13	(1	1	1	0	1	....	1)

Multivariate time series data

Chronological order

Figure 4-2 An example of multivariate time series data

4.1.2.3 Development of a GTM-TT service trend map

After data preprocessing, the GTM-TT algorithm is employed to create a GTM-TT service trend map for identifying service trends with multivariate time series data, that is, chronological keyword vectors. In this study, a proposed GTM-TT service trend map is defined as a two-dimensional map using GTM-TT for identifying service trends according to the change of time and consists of two maps; a service cluster map and a service sequence map. Where, a *service cluster map* represents the kind of services that are being developed and launched during that time, and service clusters are expressed by square nodes on the two-dimensional map. And a *service sequence map* exhibits the flow of service changes by using several arcs automatically provided by GTM-TT. The main components of a GTM-TT service trend map are as follows:

- Node: cluster of similar services in similar time
  - Node size: scaled according to the ratio of clustered services

- Arc: linkage between service clusters according to the change of time, which can represent service dynamics between service clusters
  - Line: service change path
  - Square: start point of service change
  - Circle: end point of service change

Once nodes, which represent service clusters, are identified on the two-dimensional map, arcs starting from square to circle among nodes, which represent the dynamic change path of services, are also provided by GTM-TT. An example of a GTM-TT service trend map is depicted in Figure 4-3. From this example, service clusters of a specific service field are explained by four nodes including S1, S2, S3, and S4 on the service cluster map, and these service clusters are changed from S1 to S4 in order, according to the line starting from square to circle on the service sequence map. As shown in Figure 4-3, the service trend is identified intuitively through the proposed GTM-TT service trend map.

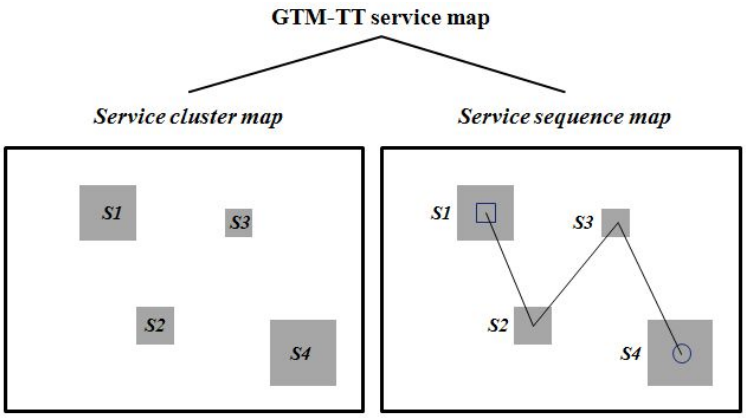


Figure 4-3 Structure of a GTM-TT service trend map

As a result of employing GTM-TT, services are clustered, and they are displayed with a trajectory on a rectangular planar surface. In this research, MATLAB software is used to employ GTM-TT. In a GTM-TT service trend map, clusters are represented by squares that are scaled according to the ratio of data points and linked by contextual information for each data point. Unlike other clustering models including standard GTM, different series are represented in the latent space by different distributions and trajectories when GTM-TT is employed since GTM-TT considers contextual information for each data point. This indicates that the model is actually identifying the difference between series. Therefore, GTM-TT algorithm is used to identify the service trends.

4.1.2.4 Interpretation

According to the result of investigating all services in the clusters, a GTM-

TT service trend map (Figure 4-3) obtained by GTM-TT algorithm is interpreted from a trend analysis view. In terms of the service cluster, the titles, categories, descriptions, etc of services in the service cluster are investigated to identify the characteristics of the obtained clusters and label them. In terms of the dynamic change path of the derived service clusters according to the change of time, the arc on the service map is utilized.

4.1.3 Advantage of the Proposed Approach

The main advantages of the proposed approach are as follows. The first stems from the use of computer-based techniques in order to deal with the vast amount of service data, which is becoming increasingly difficult to explore and analyze. The proposed method uses the developed JAVA programming to collect data and construct the database, and additionally, when extracting keywords, the text-mining technique is used. This method can save the time and efforts of researchers. Next, since the GTM-TT automatically provides not only service clusters on a two-dimensional map, but also sequence between service clusters for identifying service trends, researchers can understand service trends, intuitively. In sum, the visualization and automation of the proposed approach is the main advantage.

4.1.4 Case Study: Camera Technology-Based Mobile Application Service

A case study of camera technology-based mobile application service is presented to illustrate the suggested approach. We considered this case example appropriate for the following reasons. First, since the mobile application service is considered as a kind of service, they can be used as service data. Particularly, mobile application services are a specific subset of electronic services, which are defined as services that are offered via mobile and wireless networks (Bouwman et al., 2008).

Second, camera technology-based mobile application service is considered as one of the most critical mobile application services because mobile application services based on the mobile camera technology have been occupied a large portion of total mobile application services, and camera technology-based mobile application services have been used for various purposes, such as taking a picture, scanning barcodes, video recording, and video calling. Third, the various advanced application services using mobile camera technology are continuously introduced, and the amount of customer needs are increasing, and therefore, trend analysis of this case is useful. Finally, the number of service documents in this case is a convenient size for illustrating the proposed approach.

4.1.4.1 Construction of database

The AppStore served as the data source for collection of mobile application service documents. A total of 432 service documents about camera technology-based mobile application service were collected over the period

of 2010 first with the aid of a JAVA program, which we developed for this case, and then, with the help of the domain experts. Among all collected service documents, top 100 service documents with high average rate were used to consider the quality of services. The constructed database included a variety of information, such as the name of the mobile application service, last changed date, category, developer, version, size, launching date, and description, as shown in Table 4-2. In this study, the categories title of mobile application service, average rating, first launching date, and description of mobile application service were employed to develop the GTM-TT service trend map.

Table 4-1 Constructed database

Title	Last Changed Date	Category	Developer	Version	Average Rating (reviews)	Size(mb)	Launching Date	Description
PhotoCalc	10/7/2	Photography	Adair systems, LCC	1.2.3	3.00(421)	0.6	08/8/27	Photocal is a utility…
Camera Art	10/8/14	Photography	Sudobility	3.1.5	2.50(39)	1.1	08/12/2	This app is ispired…
Spy remote	11/1/6	Utilities	Flying pig	1.2.1	3.00(39)	0.1	09/3/8	Spy remote allows you…
				. . .				
ABABasic	11/5/3	Education	KV Adaptive LLC	1.2	2.00(8)	0.1	09/11/18	Designed by a BCBA and biomedical engineer…

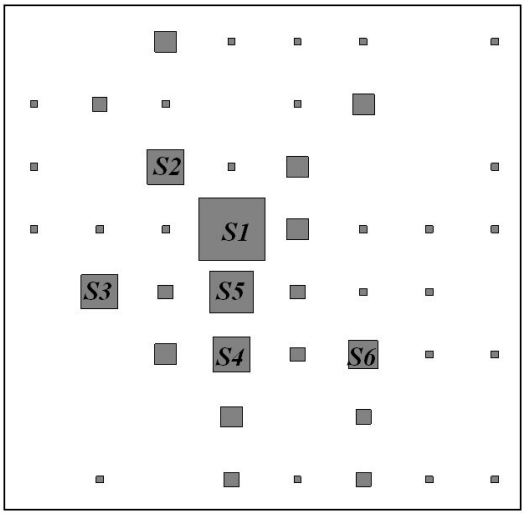
4.1.4.2 Data preprocessing

The output of keyword extraction for camera technology-based mobile application service documents is the keyword list. For this, A TextAnalysis 2.32 software was first used as a text-mining tool. After executing text-mining, the keywords are listed in the order of frequency, and keywords with high frequency are considered in the analysis based on the assumption that more important keywords appear more frequently. Finally, as a result of the keyword screening, through which the domain experts eliminate meaningless or redundant keywords, 26 keywords remained, and these will be the basis of the following analysis.

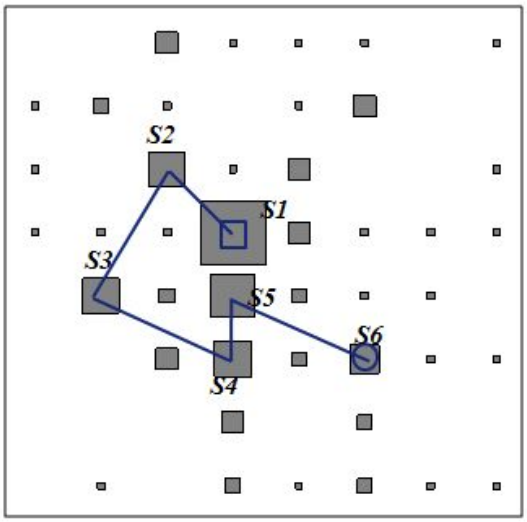
With the extracted 25 keywords including portrait, video, augmented reality, resize and so on shown in Appendix D, 100 keyword vectors for individual service documents were constructed. Note that the fields in *keyword vector* are filled with the occurrence, which is expressed with 0 or 1. For example, the occurrence of 1 for “portrait” field in the service document 1 means that “portrait” occurs in the document of service 1. Finally, all keyword vectors are arranged in the order of launching date to consider contextual information in the analysis. In other words, multivariate time series data to be used as input data for developing a GTM-TT service trend map are constructed.

4.1.4.3 Development of a GTM-TT service trend map

As pointed out before, the GTM-TT service trend map was developed by executing the GTM-TT with the constructed multivariate time series data. The GTM-TT service trend map consists of two maps including a service cluster map and a service sequence map, as shown in Figure 4-4. In practice, a GTM-TT service trend map is constructed differently according to the parameters used for employing GTM-TT. In other words, the shape of the GTM-TT service trend map is affects the number and features of service clusters. And the shape of the GTM-TT service trend map is changes by different parameters, such as the number of grids, the basis function, the number of iterations, and so on.



(a) Service cluster map



(b) Service sequence map

Figure 4-4 GTM-TT service trend map for camera technology-based mobile application service in 2010

4.1.4.4 Interpretation for the service cluster map

In the Figure 4-4(a), the top 6 total of service clusters, including S1, S2, S3, S4, S5, and S6, are selected as service clusters of camera technology-based mobile application services in 2010. For each cluster, specifically, the titles, categories, descriptions, and keywords that have been indexed in the text mining process are reinvestigated to label the obtained service clusters. As a result, S1 is labelled as “Entertainment service” since this service cluster mainly consists of mobile application services for editing photos for fun and has keywords including fun, resize, draw, and text. S2 is named as “Location information-based service” because many of mobile application services in

this service cluster are representing location information on the photos or maps and location, text, fun, and so on are included as keywords. S3 is a "Video related service" which has mobile application services for taking, editing, managing, or analyzing a video and includes keywords such as video, movie, text, etc. S4 is named as "Social communication service" which mostly comprises mobile application services to be used for distributing photos and videos among friends and acquaintances and description and keywords of mobile application services in S4 are highly related to the social communication. In a similar way, S5 and S6 are labelled as "Professional purpose service" and "Healthcare service", respectively. The result is summarized as shown in Table 4-3 with three examples of mobile application services. In sum, each individual service cluster has its unique characteristics, and each one is distinguished from the others.

Table 4-2 Service clusters of camera technology-based mobile application services in 2010

Group	Characteristic	Description	Mobile application services
S1	Entertainment	Editing photos for fun	Picture3D, Light Painting Camera, iPhone Self-Timer
S2	Location information-based	Expressing location information on the photos or maps	Snap+Map, Photo Locations, Duplicam Camera Module
S3	Video related	Including camcorder feature, managing private video, and editing videos	Swing Pro, Video Call for iPhone, SketchPad Pro
S4	Social communication	Sharing photos and videos with other people	PhotoSpread, Photo Share Pro, iMotion Flashlight Pro For iPhone 4
S5	Professional purpose	Editing photos professionally	StudioVisit, True NightVision, Empire ISIS AR
S6	Healthcare service	Managing body line and personal health	Instant Heart Rate, Heart Fitness, Bodybuilding.com

4.1.4.5 Interpretation for the service sequence map

In the Figure 4-4(b), the dynamic change path of service clusters, in which S1, S2, S3, S4, S5, and S6 are connected in order, is identified. At the beginning of 2010, mobile application services related to *entertainment* for editing photos for fun were launched primarily. Then mobile application services based on *location information* using mounted GPS on the smartphone were released in order to satisfy the needs of smartphone users want to know the location information. After that, the service trend was changed to *video-related* mobile application services because of the needs of customers for video. And then, mobile application services for *social communication* sharing original or edited photos and videos with other people such as friends and family were provided. The next is mobile application services for *professional purpose* to edit photos for specific purposes professionally. And mobile application services providing *healthcare service* were launched.

4.1.5 Discussion

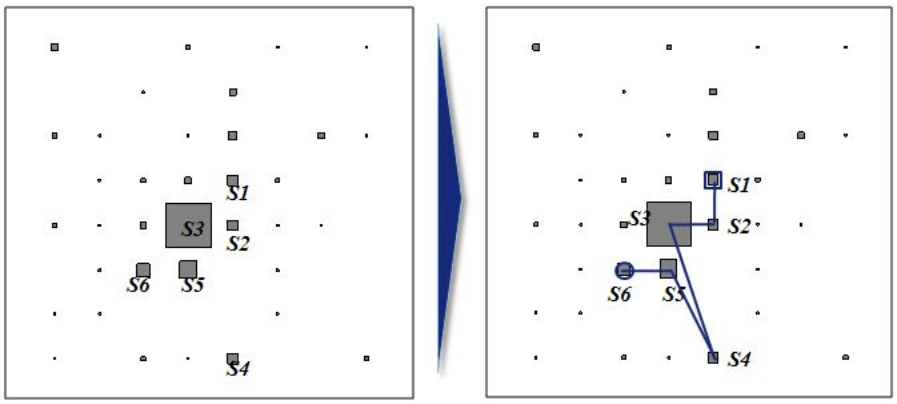
4.1.5.1 Dynamic analysis

Since this map provides the dynamic trends of services, setting an appropriate period determines the quality of analysis. The determination of period solely depends on the analyst's decision, according to the purpose of

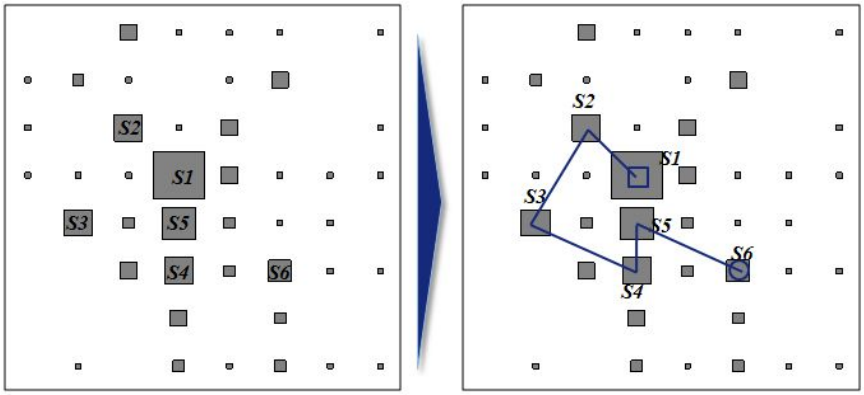
analysis. For example, if a firm wants to investigate the general pattern of three-year service trends, the scope of analysis is determined as three periods. If a firm wants to analyze the period-wise trends, the scope of analysis can be a single year with three different maps.

To compare the results, we conducted additional analysis. On one hand, we constructed three dynamic maps for 2009, 2010, and 2011. A total of 300 service documents (100 for each period) with high average rate score were collected and pre-processed. And then, the GTM-TT service trend map for each year is constructed using GTM-TT. To compare the result, we also analyzed the service trend for entire three years, using 300 service documents for 3-year period (from 2009 to 2011) as well. For both analyses, 8 × 8 squared grids were used to train both models.

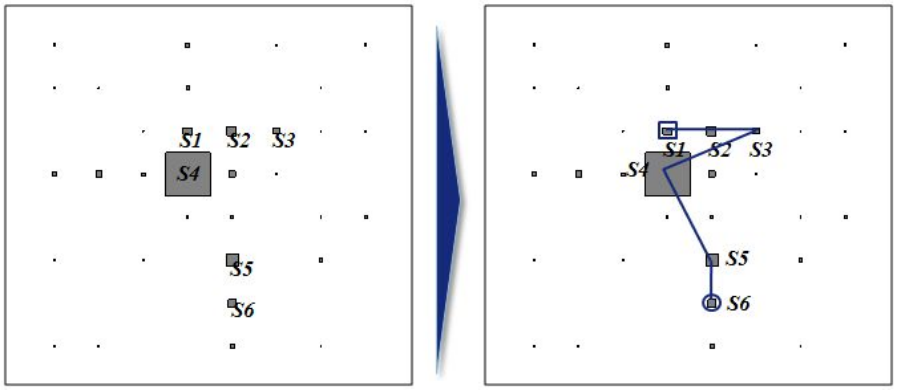
As a result, Figure 4-5 shows the GTM-TT service trend map for 2009, 2010, and 2011, respectively.



(a) Service cluster map (left), service sequence map (right) for 2009



(b) Service cluster map (left), service sequence map (right) for 2010



(c) Service cluster map (left), service sequence map (right) for 2011

Figure 4-5 GTM-TT service trend map for each year (2009, 2010, and 2011)

The service trends for 2009, 2010, and 2011 are described as shown in Table 4-4. For all three periods, a service cluster for entertainment is shown as the first cluster that shapes the overall trends for each period. Some differentiated patterns are also found according to the period.

1) Service trends for 2009

For example, in 2009, two main themes of service are found: entertainment

and social communication services. These services dominate the service trends, appearing and disappearing according to the timeline. Services such as Sports Brush, iJuly4th, and How2Draw Faces are the representative services in this period. All are the services to provide entertainment for uses, by providing a paint brush or providing unique digital pictures. The second cluster, communication cluster shows the communication-based services such as Reflections, Visual Fusion Contribute, and Emoji Brush Lite. Most of these provide additional communication with their friends after taking pictures. The third cluster, entertainment service shows the highest group among the service trend. As well, services with professional purposes such as PhotoCanvas, Foto Flair, and Loupe are found in the next trend. Most of these are related to the photo-editing services, importing services, or creative artworks. Then, services for social communication and entertainment are followed. These service trends are summarized in Table 4-4.

Table 4-3 Service trends of camera technology-based mobile application services in 2009

Group	Characteristic	Description	Mobile application services
S1	Entertainment	Creating different version of pictures or customizing photos	Sports Brush, iJuly4th, How2Draw Faces
S2	Social communication	Sharing experiences or emotions	Reflections, Visual Fusion Contribute, Emoji Brush Lite
S3	Entertainment	Adding a new function to the photos	Bobble Buddy, Zapp, HoloSnaps, SmackShot Lite, Wiggle3D
S4	Professional purpose	Providing feature-rich and easy-to-use photo editing/enhancing tool	PhotoCanvas, Foto Flair, Loupe
S5	Social communication	Providing easy update or easy access to the social network	Pichirp Pro, QUICKPING - Update 40 Social Networks At Once, Update Tycoon - Facebook + Twitter Tool
S6	Entertainment	Adding functions for fun	VooDude, Photo Geo, Tic Tac Face

2) Service trends for 2010

The service trends have been slightly changed in 2010, showing more specific patterns such as location-based service, video-related services, or professional services. Starting from the entertainment services which take a big seat in the entire mobile services, the service trend has been changed to the location based services, video-related services, social communication services, professional services, and healthcare services, as shown in Table 4-3. For more information, please refer to the previous section to describe the service trends for 2010.

3) Service trends for 2011

In 2011, the fourth cluster, entertainment also plays a critical role for shaping the service trend. The trend has been changed from entertainment, utility-based service, video-related service, and back to the entertainment services. Compared to the previous periods, services related to the technology such as utility or video-related services show the dynamic trends in 2011. However, similar to other periods, entertainment services still occupy a big seat in service trends.

As shown in Table 4-5, in 2011, utility services such as iCarBlackBox Lite, SnapTimer Lite, NO Camera+, and Private Camera 2 provide the utilitarian value for customers using camera function. For example, iCarBlackBox provides evidences in a case of car-accident, by automatically recording a car-accident. The next trend is related to the video-related services such as SynthCam, iSentry, and Portable Monitor.

Extended from taking pictures, these services provide the video-recording services or motion-sensing function to the customers. Similar to other period, entertainment service and social communication service are also frequently found in 2011 as well.

Table 4-4 Service trends of camera technology-based mobile application services in 2011

Group	Characteristic	Description	Mobile application services
S1	Entertainment	Providing additional functions for specialized experiences	Corel Paint it! Now, Happy Chinese New Year 2012 Dragon Photo Fun, Pacific Park on the Santa Monica Pier
S2	Utility	Providing special functions such as automated recoding, protection, and privacy	iCarBlackBox Lite, SnapTimer Lite, NO Camera+, Private Camera 2 - protect your photos and videos
S3	Video related	Adding functions for video-related applications	SynthCam, iSentry, Portable Monitor (Camera to PC withOUT Client)--Third Eye Lite
S4	Entertainment	Providing fun by customizing the pictures such as color change, browsing, and retouching	Makeup Simulator, La Koketa - Your digital wardrobe and modern go-to stylist, ColorRetouch
S5	Social communication	Sharing photos or information with friends or family	iRate Bikez Pro, iRate Tats Pro, Paul
S6	Entertainment	Customizing pictures or their own backgrounds	Dog Booth, Glow Wallpapers for iPhone!, FaceBooth for iPhone Free

4.1.5.2 Period determination of GTM-TT service trend map

Since the aim of this study is to construct the GTM-TT service trend map to identify the trends of service, dynamically, period determination is a significant issue to be considered. For example, the GTM-TT service trend map can be constructed for various time periods.

For example, three snapshot of Figure 4-5 which shows the yearly trends can be differently constructed. For investigating three-year dynamics, we can gather the service data for three years and construct a single service map, illustrating the entire service trend for three years in a single map, as shown in Figure 4-6. The result of three-year analysis is shown in Table 4-6.

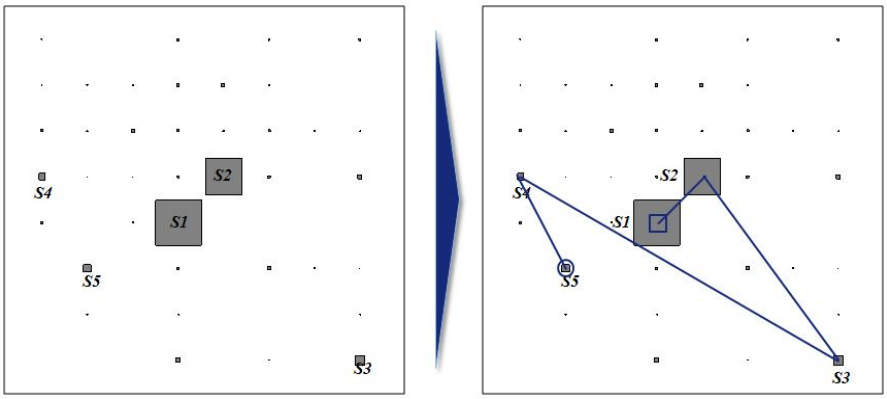


Figure 4-6 Service cluster map (left), service sequence map (right) (from 2009 to 2011)

Table 4-5 Derived service trends for three years (from 2009 to 2011)

Group	Characteristic	Description	Mobile application services
S1	Entertainment	Creating different version of pictures or customizing photos	Sports Brush, Bobble Buddy, iJuly4th
S2	Utility	Providing useful functions by integrating camera	Smart Counter, PhotoFTP5, Alford Calc
S3	Social communication	Sharing photos or information with friends or family to survey	iRate Dogs Pro, iRate Bunnies Pro, iRate Tats Pro
S4	Photography	Providing skills and tips for photography	Photography Tips from QuickPro, Viewfinder Pro, Digital 201 from QuickPro
S5	Utility	Providing functions such as timer, thermal vision effect	SnapTimer Lite, Camera Timer+ Lite, Thermal infrared camera

Comparing the result of Figure 4-5 and Figure 4-6, we can derive the important implication for determining the period for constructing the GTM-TT service trend map. Compared to the service map of three-year period, that of one-year period shows the close link between clusters. In other words, the distance between clusters seems to be closer in one-year period. This means that each cluster in three-year period is more unrelated and distinct than that in one-year period. This result is natural, since only 5 clusters are found in entire three-year period in Figure 4-6 whereas 6 clusters are found in each year in Figure 4-5. Of course, service trends in three-year period might be more generic whereas that in one-year period seems to be more specific. Summarizing the result, Table 4-7 summarizes the result of comparison different periods.

Table 4-6 Comparison of period determination

Comparison	Shorter period (e.g. 1 year)	Longer period (e.g. 3 years)
Purpose	Analyzing specific patterns	Analyzing general patterns
Distance of clusters	Relatively close	Relatively distant
Result	Specific	General

The determination of period can be decided according to the analyst's decision. If one wants to see the general patterns of longer period, a single map can provide a clear and neat view. However, if a firm wants to see more specific results, analysts should set the analysis period short, in order to provide more detailed view.

However, even when we determine the analysis period as longer, specific patterns can be also derived by controlling the number of grids. For example, currently 8 × 8 grids are used to provide the result. If more detailed information is required, one can increase the number of grids in order to derive more specific and detailed clusters. It should be noted that the generalization or specification of result does not solely depends on period determination.

Chapter 5. Conclusions

5.1 Summary and contributions

The dissertation proposed the systematic approaches to aiding identification of new opportunities and analysis of trends in technology or service through developed GTM-based maps using information visualization techniques. It is shown that GTM and GTM-TT are viable and effective for developing and applying GTM-based maps. Both are also promising methodologies for another critical problem that has not been dealt with. However, proposed approaches do not actually 'solve' problems; that is, they do not provide global optimum solutions. What they do is to 'aid' designers to understand the big picture under a given context

In summary, the first study proposed a framework to develop a GTM-based patent vacuum map for identifying new technology opportunities. Patents related to lithography technology were utilized as data to represent technology. In this study, a total of 13 patent vacuums were identified as new technology opportunities since it means that these technologies have not developed or launched so far. As well, comparing with the SOM-based patent map and the PCA-based patent map was conducted for validating and showing the usefulness as another alternative of proposed the GTM-based patent vacuum map.

The second study developed a GTM-based service vacuum map and applied it to the search stage of NSD using Apple AppStore data. GTM is employed to take existing services and service vacuums into account for the purpose of identifying new service opportunities. The proposed GTM-based service vacuum map provides insights into new service opportunities by identifying and visualizing unexplored vacuums in the map. These are blank grids that are surrounded by plentiful existing services that can be expected to offer potential as new services. Promising new service opportunities can be derived by interpreting the identified service vacuums through the inverse mapping function of GTM. And the characteristics of service vacuums were provided through reinvestigation of derived vacuums.

Note that the derived all patent and service vacuums from the developed GTM-based vacuum maps do not indicate innovative technologies and services. The results only represent potentials to be new technologies and services. Therefore, above two studies relate to incremental innovation rather than radical innovation.

The third study suggested a GTM-TT service trend map as a technology map for identifying trends of service. In this study, the detailed process of the proposed approach is provided, and a case study of camera technology-based mobile application service is conducted to illustrate the proposed approach. Specifically, a JAVA program is first developed for collecting data from a data source and constructing a database. Multivariate time series data are then constructed with the aid of text-mining technique and domain experts. Thirdly, a GTM-TT service trend map is developed by

implementing a GTM-TT algorithm. Finally, interpretation of the resulted GTM-TT service trend map including a service clustering map and a service sequence map is conducted.

The contribution of the dissertation is to propose the systematic approaches to developing and applying GTM-based maps using information visualization that have not been dealt with intensively so far, or do not have satisfactory solutions, but are significantly important for successful implementing information visualization in technology or service. In practice, the proposed approaches are expected to aid designers to implement information visualization. In terms of theoretical developments, the dissertation further extended the applicability of the GTM and GTM-TT.

More specifically, each study has its unique contributions, as stated in the three sub-conclusions. The first study has following contributions. Firstly, the intuitive and objective detection of patent vacuums compared to the previous techniques is a result of the GTM-based patent vacuum map, achieved by the grid-based visualization algorithm of the GTM. The second contribution comes from the automatic and objective interpretation of patent vacuums due to the inverse mapping function of GTM, which enables the transformation of the latent variable into the original data space. Thus, researchers, engineers and managers interested in new technology development save time and energy when uncovering new technology opportunities as well as acquiring objective results.

The main contributions of the second study are threefold. First, this study theoretically contributes to NSD research proposing an intelligent

approach. Second, this study is exploratory in that the inverse mapping of a GTM model is first proposed. Hence, this research emphasizes on how to interpret service vacuums using the inverse mapping and its strengths for identifying new service opportunities. Lastly, since all activities in the process could be computerized, researchers, engineers, managers, etc., who are interested in NSD can save considerable time and effort when uncovering New service opportunities in terms of managerial implications.

The third study proposed an approach which can be employed for dynamic service trend analysis, ranging from the identification of service trends to the monitoring dynamic change path of service trends. Next, it is possible to understand the service trends at the service level as opposed to the company or industry level, which could be applicable to various research domains. Finally, this is an intelligent and pioneering approach that can analyze and visualize service trends and the dynamic change path of them, unlike previous qualitative approaches.

5.2 Limitations and future research

The dissertation is subject to some limitations, which could serve as fruitful avenues for future research.

On the whole, firstly, a necessity in elaborating the keyword extraction process is vital, since keyword extraction plays a critical role in determining the value of analysis results. Although the text mining tool and expert judgment were employed to extract the keywords, covering both

quantitative and qualitative perspectives of keyword extraction, other systematic methodologies should supplement these techniques in order to validate the extracted keywords.

Secondly, both GTM and GTM-TT are very sensitive to parameter settings potentially resulting in inappropriate map if the parameters are set incorrectly. A systematic guideline for defining parameters is potentially another subject matter requiring further study.

Thirdly, the validity of this approach necessitates more testing work employing other data, which will be indispensable for gaining external validity. Especially, the test whether collected multivariate data can be mapped on the two-dimensional map is needed.

Fourthly, the third study is conducted for monitoring purpose. However, the importance of forecasting future trend is increasing in terms of innovation of technology or service. So, future research, which is elaborating this study to be utilized for forecasting purpose, would be valuable.

Finally, the whole process needs to be systemized and automated. Although an automated supporting system has been developed, there is still considerable scope for further work to enhance operational efficiency. Especially, the automated system for collecting dynamic data or real time data is necessary to make proposed approaches more useful. These topics would be fruitful areas for future research.

Bibliography

Andrade, A., Nasuto, S., Kyberd, P., and Sweeney-Reed, C.M. (2005). Generative topographic mapping applied to clustering and visualization of motor unit action potentials. *BioSystems*, 82(3), 273-284.

Arbnor I. and Bjerke B. (1997). *Methodology for creating businesses knowledge*. Thousand Oaks, CA/London: Sage, 548.

Bao, L., Huang, Y., and Song, J. (2009). On the Common Information Service Platform Based Supply Chain Management, a Case Study, Information Engineering and Electronic Commerce, 740-744.

Basberg, B.L. (1987). Patents and the Measurement of Technological Change: A Survey of Literature. *Research Policy*, 16(2/4), 131-141.

Benbunan-Fich, R. and Benbunan, A. (2007). Understanding user behavior with new mobile applications. *Journal of Strategic Information Systems*, 16(4), 393-412.

Bianchini D., De Antonellis V., and Melchiori M. (2006). Lightweight Ontology-Based Service Discovery in Mobile Environments, Database and Expert Systems Applications, 359-364.

Bishop, C.M., Hinton, G.E., and Strachan, I.G.D. (1997). GTM Through Time. *In Proceedings 1997 International Conference on Artificial Neural Networks*, 111-116.

Bishop, C.M., Svensén, M., and Williams, C. (1996). GTM: a principled alternative to the Self-Organizing Map. *In Proceedings 1996*

International Conference on Artificial Neural Networks, ICANN'96, 165-170.

Bishop, C.M., Svensén, M., and Williams, C. (1998). Developments of the generative topographic mapping. *Neurocomputing*, 21(1), 203–224.

Bouwman, H., De Vos, H., and Haaker, T. (2008). *Mobile Service Innovation and Business Models*. Berlin: Springer.

Bullen, R.J., Cornford, D., and Nabney, I.T. (2003). Outlier detection in scatterometer data: Neural network approaches. *Neural Networks*, 16(3-4), 419-426.

Card, S., Mackinlay, J., and Shneiderman, B. (1999). *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers Inc., San Francisco, CA.

Carreira-Perpiñan, M. A. (2000). Reconstruction of sequential data with probabilistic models and continuity constraints. In S. Solla, T. Leen, and K. R. Muller (Eds.), *Advances in neural information processing systems 12*, NIPS'1999, 414–420.

Chen, C. (2003). Information visualization versus the semantic web, In: Geroimenko, V. and Chen, C. (Eds), *Visualizing the Semantic Web: XML-based Internet and Information Visualization, Springer*, London, 15-35.

Chen, C. (2010). Information visualization. Wiley Interdisciplinary Reviews: Computational Statistics, 2(4), 387-403.

Choi C., Kim S., and Park Y. (2009). A patent-based cross impact analysis for quantitative estimation of technological impact: The case of

information and communication technology, *Technological forecasting and social change*, 74(8), 1296-1314.

Choi, C. and Park, Y. (2009). Monitoring the organic structure of technology based on the patent development paths. *Technological Forecasting & Social Change*, 76(6), 754–768.

Collier D. and Meyer S. (1998). A service positioning matrix, *International journal of operations & production management*, 18(12), 1223-1244.

del-Hoyo R., Martin-del-Brio B. and Medrano N. (2009). Computational intelligence tools for next generation quality of service management, *Neurocomputing*, 72(16/18), 3631-3639.

Delen, D. and Crossland, M.D. (2008). Seeding the survey and analysis of research literature with text mining. *Expert systems with applications*, 34(3), 1707–1720.

Eppler, M.J. and Burkhard, R.A. (2007). Visual representation in knowledge management: framework and cases. *Journal of Knowledge Management*, 11(4), 112-122.

Ercan, G. and Cicekli, I., (2007). Using lexical chains for keyword extraction. *Information processing & management*, 43(6), 1705-1714.

Eric, S. and Sergios, D. (2005). Managing the new service development process: towards a systemic model. *European Journal of Marketing*, 39(1/2), 175-198.

Ernst, H. (2003). Patent information for strategic technology management. *World Patent Information*, 25(3), 233–242.

Faia-Correia, M., Patriotta, G., Brigham, M., and Corbett, J.M. (1999).

Making Sense of Telebanking Information Systems: The Role of Organizational Back-Ups. *Journal of Strategic Information System,* 8(2), 143–56.

Fay, B. (2002). Advanced optical lithography development from UV to EUV. *Microelectronic Engineering*, 61/62, 11-24.

Feldman, R., Fresko, M., Hirsh, H., Aumann, Y., Lipshstat, O., Schler, Y., and Rahman, M. (1998). Knowledge management: a text mining approach, *Proceeding of the 2nd International Conference on Practical Aspects of Knowledge Management*, 9-10.

Fisk, R.P., Brown, S.W., and Bitner M.J. (1993). Tracking the evolution of the services marketing literature. *Journal of retailing*, 69(1), 61-103.

Ganapathy, S., Ranganathan, C., and Sankaranarayanan, B. (2004). Visualization strategies and tools for enhancing customer relationship management. *Communication of ACM*, 47(11), 92–99.

Gebauer, H. and Friedli, T. (2005) Behavioral Implications of the Transition Process from Products to Services. *Journal of Business and Industrial Marketing*. 20(2), 70-78.

Gilson, O., Silva, N., and Grant, P.W. (2008). From Web Data to Visualization via Ontology Mapping, *Journal of the European Association for Computer Graphics*, 27(3), 959-966.

Girolami, M. (2002). Latent variable models for the topographic organization of discrete and strictly positive data. *Neurocomputing*, 48(1-4), 185–198.

Grandstrand, O. (1999). *The Economics and Management of Intellectual Property: Toward Intellectual Capitalism*. Cheltenham: Edward Elgar.

Grilliches, Z. (1990). Patent Statistics as economic indicators: A Survey. *Journal of economic literature*, 28(4), 1661-1707.

Hara T., Arai T. and Shimomura Y. (2009). A Method to Analyse PSS from the Viewpoints of Function, Service Activity & Product Behaviour, *Industrial product-service systems*, 180-185.

Harriott, L.R. (2001). Limits of Lithography. *Proceedings of the IEEE*, 89(3), 366-374.

Hofacker, C.F., Goldsmith, R.E., Bridges, E., and Swilley, E. (2007). *E-services: a synthesis and research agenda*. New York, DUV, 13-44.

Hogo, M. (2010). Evaluation of e-learning systems based on fuzzy clustering models and statistical tools. *Expert Systems with Applications*, 37(10), 6891-6903.

Holdford D. and Kennedy D. (1999). The service blueprint as a tool for designing innovative pharmaceutical services, *Journal of the American Pharmaceutical Association*, 39(4), 545-552.

Jaffe, A.B., Trajtenberg, M., and Forgarty, M.S. (2000). Knowledge Spillovers and Patents Citations: Evidence from a Survey from Inventors. *American Economic Review*, 90(2), 215-218

Janakiram, M. and Goernitz, S. (2005). Real-Time Lithography Registration, Exposure, and Focus Control—A Framework for Success. *IEEE transactions on semiconductor manufacturing*, 18(4), 534-538.

Jaw, C., Lo, J.Y., and Lin, Y.H. (2010). The determinants of new service development: service characteristics, market orientation, and actualizing innovation effort. *Technovation*, 30(4), 265-277.

Johnson, R. and Wichern, D. (1988). *Applied multivariate statistical analysis*. New Jersey: Prentice Hall.

Karki M. (1997). Patent Citation Analysis: A Policy Analysis Tool, *World patent information*, 19(4), 269-272.

Keim, D. (2001). Visual exploration of large data sets. *Communications of the ACM*, 44(8), 38–44.

Keim, D., Andrienko, G., Fekete, J., Görg, C., Kohlhammer, J., and Melançon, G. (2008). *Information Visualization*. Lectures Notes in Computer Science, Springer Berlin / Heidelberg, 154–175.

Keller, T. and Tergan, S. (2005). Visualizing knowledge and information: an introduction, In Tergan, S. and Keller, T. (Eds), *Knowledge and Information Visualization*, Spirnger, Berlin, New York, 1-23.

Kim, C., Choe, S., Choi, C., and Park, Y. (2008). A systematic approach to new mobile service creation. *Expert Systems with Applications*, 35(3), 762-771.

Kim, J. and Park, Y. (2010). Identifying a New Service Opportunity from Potential Needs: User-centric Service Map. *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*. 357-361.

Kim, Y.G., Suh, J.H., and Park, S.C. (2008). Visualization of patent analysis for emerging technology. *Expert Systems with Application*, 34(3), 1984-1812

Kingman-Brundage J. (1994). Service Mapping: Gaining a Concept Perspective on Service System Design, *QUIS*, 3, 235-246.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1), 59-69.

Kohonen, T. (1995). *Self-organizing maps*, Berlin: Springer.

Kwak, R., Lee, H., and Park, Y. (2010). On the construction of a service map: How to match the service features and the customer needs. *Networking and Information Technology (ICNIT), International Conference on*, 298-302.

Lamirel, J.-C. and Al Shehabi, S., (2006). MultiSOM: A Multiview Neural Model for Accurately Analyzing and Mining Complex Data. *Fourth International Conference on Coordinated & Multiple Views in Exploratory Visualization (CMV)*.

Lee, C., Jeon, J., and Park, Y. (2011). Monitoring trends of technological changes based on the dynamic patent lattice: a modified formal concept analysis approach. *Technological Forecasting and Social Changes,* 78 (4), 690-702.

Lee, S., Yoon, B., and Park, Y. (2009). An approach to discovering new technology opportunities: Keyword-based patent map approach. *Technovation*, 29(6-7), 481-497.

Levinson, H. (2001). *Principles of Lithography*. Bellingham, WA: SPIE, 159.

Li, Y., Wang, L., and Hong, C. (2009). Extracting the significant-rare keywords for patent analysis. *Expert Systems with Applications*, 36(3), 5200-5204.

Liang, Y. and Tan, R. (2007). A text-mining-based patent analysis in product innovative process. *Trends in computer aided innovation*, 250, 89–96.

Lin, H.H. (2011). The effect of multi-channel service quality on mobile customer loyalty in an online-and-mobile retail context. *The Service Industries Journal*, iFirst Article, 1–18.

LogicaCMG Press Release, (2005). *Mobile content market set to triple to more than 7.6 billion euros within a year*, July 6, <http://www.logicacmg.com/pressroom>.

Losiewicz, P., Oard, D., and Kostoff, R. (2000). Textual data mining to support science and technology management. *Journal of Intellectual Information System,* 15(2), 99-119.

Manning, C.D., Raghavan, P., and Schutze, H. (2008). *Introduction to Information Retrieval*. Cambridge Univ. Press, Cambridge.

Martino, B. (1983). An equation of growth of a single species with realistic dependence on crowding and seasonal factors, *Journal of mathematical biology*, 17(1), 33-43.

Martino, J.P. (2003). A review of selected recent advantages in technological forecasting. *Technological Forecasting & Social Change*, 70(8), 719-733.

McKim, R.H., (1980). *Experiences in Visual Thinking*, Boston, MA: PWS Engineering.

Menor, L.J. and Roth, A.V. (2008). New Service Development Competence and Performance: An Empirical Investigation in Retail Banking. *Production and Operations Management*, 17(3), 267–284.

Meyer, J., Thomas ,J., Diehl, S., Fisher, B., Keim, D., Laidlaw, D., Miksch, S., Mueller, K., Ribarsky, W., Preim, B., and Ynnerman, A. (2007).

From visualization to visually enabled reasoning. Technical report, Dagstuhl Seminar 07291 on Scientific Visualization.

Moens, M. (2006). *Information Extraction: Algorithms and Prospects in a Retrieval Context*. Springer, Dordrecht.

Morelli, N. (2002). Designing Product/Service Systems: A Methodological Exploration, *Design issues*, 18(3),.3-17.

Nieto M., Arias D., and Minguela B. (1999). The evolution of operations management contents: an analysis of the most relevant textbooks, *Industrial management and data systems*, 99(8), 345-353

Olier, I. and Vellido, A. (2008). A Variational Formulation for GTM Through Time. *IEEE International Joint Conference on 2008 June*, 516-521.

Oliva, R. and Kallenberg, R. (2003). Managing the Transition from Products to Services. *International Journal of Service Industry Management*, 14(2). 160-172.

Park, Y., Yoon, B., and Lee, S. (2005). The idiosyncrasy and dynamism of technological innovation across industries: patent citation analysis. *Technology in Society*, 27(4), 471–485.

Pérez, J.M., Berlanga, R., Aramburu, M.J., and Pedersen, T.B. (2008). Integrating data warehouses with web data: A survey. *IEEE transactions on knowledge and data engineering*, 20(7), 940-955.

Phaal R., Farrukh C., and Probert D. (2005). Developing a Technology Roadmapping System, *Portland international conference on management of engineering and technology*, 99-111.

Pilat, D. (2000). No longer services as usual. *The OECD Observer*, 223, 52–54.

Pitt, L.F., Parent, M., Junglas, I., Chan, A., and Spyropoulou, S. (2010). Integrating the smartphone into a sound environmental information systems strategy: Principles, practices and a research agenda. *Journal of Strategic Information Systems,* 20(1), 27-37.

Ray, A.W. and Ray J.J. (2006). Strategic benefits to SMEs from third party web services: An action research analysis. *Journal of Strategic Information Systems*, 15(4), 273-291.

Robertson, N. and Perera, T. (2002). Automated data collection for simulation?. *Simulation practice and theory,* 9(6/8), 349-364.

Rosenberg, D. (1997). *Trend analysis and interpretation: key concepts and methods for maternal and child health professionals*. Maternal and Child Health Information Resource Center, US Department of Health & Human Services.

Russell, A., Chiu, C., and Korde, T. (2009). Visual representation of construction management data. *Automation in Construction*, 18, 1045–1062.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1), 1–47.

Sensoy M. and Yolum P. (2007). Cooperative Evolution of Service Ontologies, *Autonomous agents and multiagent systems*, 1206-1208

Shen, X.X., Tan, K.C., and Xie, M. (2000). An integrated approach to innovative product development using Kano's model and QFD.

European Journal of Innovation Management, 3(2), 91–99.

Sheng, H., Nah, F.F.-H., and Siau, K. (2005). Strategic implications of mobile technology: A case study using Value-Focused Thinking. *Journal of Strategic Information Systems*, 14(3), 269-290.

Shneiderman, B. (1996). The eyes have it: a task by data type taxonomy for information visualization. *Proceedings of IEEE Workshop on Visual Languages*, 336-343.

Shostack, L. G. (1982). How to Design a Service, *European Journal of Marketing*, 16(1), 49-63.

Smith, B. (1998). *Optics for photolithography,* in Microlithography Science and Technology. Sheats, J. and Smith, B. (Eds.), New York: Marcel Dekker, 263–264.

Song, B., Kang, D., Yoon, B., and Park, Y. (2010). Development of Two-layered Service Evolution Map: Structure and Development Process. *Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Macao, China, 419-423.

Stevens, E. and Dimitriadis, S. (2005). Managing the new service development process: towards a systemic model. European Journal of Marketing, 39(1), 175–198.

Stulen, R.H. and Sweeney, D.W. (1999). Extreme ultraviolet lithography. *IEEE journal of quantum electronics*, 35(5), 694-699.

Suh, J.H. and Park, S.C. (2009). Service-oriented Technology Roadmap (SoTRM) using patent map for R&D strategy of service industry.

Expert Systems with Applications, 36 (2/3), 6754-6772.

Svensén, M. (1998). *GTM: the generative topographic mapping*. Ph.D. Thesis, Aston University.

Tiun, S., Abdullah, R., and Tang, E. (2001). Automatic Topic Identification Using Ontology Hierarchy. *Proceeding of the 2nd International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2001)*, Mexico City, Mexico.

Toivonen, M. and Tuominen, T. (2009). Emergence of innovations in services, *The Service Industries Journal*, 29(7), 887–902.

van de Kar E., Herder P., and Snijders A., (2004). Sales-Supporting E-services. *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, 66-75.

Van der Pijl, G.J. (1994). Measuring the Strategic Dimensions of the Quality of Information. *Journal of Strategic Information Systems*, 3(3), 179-190.

van Wijk, J. (2005). The value of visualization. *Proc of IEEE Visualization*, 79–86.

Vellido, A. (2006). Missing data imputation through GTM as a mixture of t-distributions. *Neural Networks*, 19(10), 1624–1635.

Ware, C. (2004). *Information Visualization: Perception for Design*. Morgan Kaufman, San Francisco, CA.

Weng, S. and Liu, C. (2004). Using text classification and multiple concepts to answer e-mails. *Expert Systems with Applications*, 26(4), 529-543.

Westphal, C. and Blaxton, T. (1998). *Data mining solution*. New York: Wiley.

Wu, F.S., Hsu, C.C., Lee, P.C., and Su, H.N. (2011). A systematic approach for integrated trend analysis—The case of etching. *Technological Forecasting and Social Change*, 78(3), 386-407.

Yang J. and Chung I. (2006). A Method for Automatic Generation of OWL-S Service Ontology, *International journal of information processing systems*, 2(2), 114-123

Yang Q. Zhang X. and Li J. (2011). An overlapping model for new product development process optimization using DSM method, *International Conference on Instrumentation, Measurement, Circuits and Systems,* 53-56.

Yang, C.C. (2007). A Systems Approach to Service Development in a Concurrent Engineering Environment. *The Service Industries Journal*, 27(5), 635–652.

Yang, H. and Lee, C. (2004). A text mining approach on automatic generation of web directories and hierarchies. *Expert Systems with Applications*, 27(4), 645-663.

Yang, J. and Zhang, B. (2001). Customer data mining and visualization by generative topographic mapping methods. *Proceedings International Workshop on Visual Data Mining*, 4 September, Freiburg, Germany.

Yoon, B. (2010). Strategic visualization tools for managing technological information. *Technology Analysis & Strategic Management*, 22(3), 377-397.

Yoon, B. and Park, Y. (2004). A text-mining-based patent network: Analytic tool for high-technology trend. *The Journal of High Technology*

Management Research, 15(1), 37-50.

Yoon, B. and Park, Y. (2007). Development of new technology forecasting algorithm: Hybrid approach for morphology analysis and conjoint analysis of patent information. *IEEE Transactions on Engineering Management*, 54(3), 588-599.

Yoon, B., Yoon, C., and Park, Y. (2002). On the development and application of a self-organizing feature map-based patent map. *R&D Management*, 32(4), 291-300.

Yoon, J. and Kim, K. (2011). TrendPerceptor: A property–function based technology intelligence system for identifying technology trends from patents. *Expert Systems with Applications*, 39(3), 2927-2938.

Zhang N., Yuan S., and Chen T. (2008). Latent tree models and diagnosis in traditional Chinese medicine, *Artificial intelligence in medicine*, 42(3), 229-245.

Zhang, J., Chai, K.H., and Tan, K.C. (2005). Applying TRIZ to Service Conceptual Design: An Exploratory Study. *Creativity and innovation management*, 14(1), 34-42.

Appendix

Appendix A. Generative topographic mapping: GTM

Notations (Olier and Vellido, 2008)

$\Re^L$	Latent space
$\Re^D$	Data space
$\mathbf{u}$	L-dimensional point in latent space
$\mathbf{W}$	The matrix that generates the mapping
$\mathbf{\Phi}$	Radially symmetric Gaussians in the standard model for continuous data
β	The estimated common inverse variance of the isotropic Gaussian distributions in data space
$\mathbf{x}_n$	The observed data points
$\mathbf{y_i}$	D-dimensional point in the manifold embedded in data space: the center of the ith constrained mixture component
$\boldsymbol{R}$	The matrix of responsibilities
$\mathbf{G}$	A matrix with elements $\sum_{n=1}^{N} \mathbf{R_{in}}$ in the diagonal and zeros elsewhere
$\mathbf{X}$	The observed data matrix

The neural network-inspired GTM is a nonlinear latent variable model of the manifold learning family, with sound foundations in probability theory. It performs simultaneous clustering and visualization of the observed data,

through a nonlinear and topology-preserving mapping from a visualization latent space in $\Re^L$(with L being usually 1 or 2 for visualization purposes) onto the $\Re^D$ space in which the observed data reside. The mapping that generates the embedded manifold takes the form:

$$\mathbf{y} = \mathbf{W}\boldsymbol{\Phi}(\mathbf{u}) \tag{1}$$

where $\mathbf{u}$ is an L-dimensional point in latent space, $\mathbf{W}$ is the matrix that generates the mapping, and $\boldsymbol{\Phi}$ consists of S basis functions ϕ_S (radially symmetric Gaussians in the standard model for continuous data). To achieve computational tractability, the prior distribution of $\mathbf{u}$ in latent space is constrained to form a uniform discrete grid of M centres, analogous to the layout of the SOM units, in the form:

$$\boldsymbol{p(u)} = \frac{\mathbf{1}}{\boldsymbol{M}}\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{M}} \boldsymbol{\delta(u - u_i)} \tag{2}$$

This way defined, the GTM can also be understood as a constrained mixture of Gaussians model. A density model in data space is therefore generated for each component i of the mixture, which, assuming that the observed data points $\mathbf{x}_n$ are i.i.d., leads to the definition of a complete log-likelihood in the form:

$$\boldsymbol{L_c(W, \beta|X)} = \sum_{\mathrm{n}=1}^{\mathrm{N}} \ln \mathrm{p}(\mathbf{x_n}|\mathbf{u_i}, \mathbf{W},\ \beta) \tag{3}$$

where,

$$\mathrm{p}(\mathbf{x_n}|\mathbf{u_i}, \mathbf{W},\ \beta) = \left(\frac{\boldsymbol{\beta}}{\mathbf{2}\boldsymbol{\pi}}\right)^{\mathbf{D/2}} \exp\{-\beta/2 \parallel y_i - x_n \parallel^2\} \tag{4}$$

In Eq. (4), $\mathbf{y_i} = \mathbf{W}\boldsymbol{\Phi}(\mathbf{u_i})$ is a D-dimensional point in the manifold embedded in data space: the center of the ith constrained mixture component, while β is the estimated common inverse variance of the isotropic Gaussian

distributions in data space whose centers are $\mathbf{y_i}$. The adaptive parameters of the model can be optimized using the EM algorithm within the Maximum Likelihood framework. Matrix $\mathbf{W}$ is updated as the solution of the following system of equations:

$$\mathbf{\Phi^T G_{old} \Phi W_{new}^T - \Phi^T R_{old} X = 0} \tag{5}$$

where $\mathbf{\Phi}$ is a $M \times S$ matrix with elements $\phi_S(\mathrm{u_i})$; $\boldsymbol{R}$ is the *matrix of responsibilities*, with elements $\mathbf{R_{in}} = \mathrm{p}(\mathbf{u_i}|\mathbf{x_n})$ that define the probability of the data point $\mathbf{x_n}$ being generated by the latent point $\mathbf{u_i}$; $\mathbf{G}$ is a matrix with elements $\sum_{\mathrm{n=1}}^{\mathrm{N}} \mathbf{R_{in}}$ in the diagonal and zeros elsewhere; and $\mathbf{X}$ is the observed data matrix. Finally, parameter β is updated according to:

$$(\beta^{\mathrm{new}})^{-1} = \frac{1}{\mathrm{ND}} \sum_{\mathrm{n=1}}^{\mathrm{N}} \sum_{\mathrm{i=1}}^{\mathrm{M}} \mathrm{R_{in}} \parallel \boldsymbol{y_i} - \boldsymbol{x_n} \parallel^2 . \tag{6}$$

See Bishop et al., (1998) for further details on all these calculations.

Appendix B. Generative topographic mapping through time: GTM-TT

Notations (Olier and Vellido, 2008)

$\mathbf{q_n}$	The state $\mathbf{u_i}$ at time *n*
π_i	The estimation of the initial state probabilities
p_{ij}	The state transition probabilities

Multivariate time series are not i.i.d. data and, therefore, the standard definition of the GTM summarized in the previous sub-section can only provide a rough approximation to their proper modeling. A variation on the standard model was defined as a topology-constrained Hidden Markov Model (HMM) in Bishop et al. (1997) to deal with this limitation, namely the GTM through time.

In GTM-TT, points $\mathbf{u_i}$ in latent space are considered as hidden states and temporal dependencies are captured through the coupling of these latent points. Furthermore, the emission probabilities are controlled by the GTM mixture distribution.

The probability of an observation sequence $X = \{\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_n}, \cdots, \mathbf{x_N}\}$, given a fixed state sequence $Q = \{\mathbf{q_1}, \mathbf{q_2}, \cdots, \mathbf{q_n}, \cdots, \mathbf{q_N}\}$ is defined as:

$$\mathbf{p(X|Q)} = \prod_{\mathbf{n=1}}^{\mathbf{N}} \mathbf{p}\{\mathbf{x_n}|\mathbf{q_n} = \mathbf{u_i}\}. \tag{7}$$

where $\mathbf{q_n}$ is the state $\mathbf{u_i}$ at time *n*. The likelihood of the GTM-TT is defined as the sum of probabilities of the observation sequence given of all possible state sequences, or paths:

$$p(X,\lambda) = \sum_{\mathbf{allQ}} \boldsymbol{\pi}_{\mathbf{q_1}} \prod_{\mathbf{n=1}}^{\mathbf{N-1}} \mathbf{P}_{\mathbf{q_n q_{n+1}}} \prod_{\mathbf{n=1}}^{\mathbf{N}} \mathbf{p}_{\mathbf{q_n}}(\mathbf{x_n}), \tag{8}$$

where $\boldsymbol{\pi}_{\mathbf{q_1}}$ defines the initial state probability of Q; $\mathbf{P}_{\mathbf{q_n q_{n+1}}} = P\left[\mathbf{q_{n+1}} = \mathbf{u_j} \middle| \mathbf{q_n} = \mathbf{u_i}\right]$ is the probability of transition from one hidden state to another (and therefore captures the temporal dependencies); $p_{q_n}(\mathbf{x_n}) = p(\mathbf{x_n} \mid \mathbf{q_n} = \mathbf{u_i})$ is the probability found in Eq. (4); and $\lambda = \{\{\pi_i\}, \{P_{ij}\}, \mathbf{W}, \beta\}$ is the set of model parameters.

As in HMM, the likelihood defined above can be efficiently calculated using the forward-backward procedure. The probability of being in the state $\mathbf{u_i}$ at time n, given the observation sequence and the model, also known as responsibility $\mathbf{R}_{\mathrm{in}}$ is calculated as:

$$R_{in} = P(\mathbf{q_n} = \mathbf{u_i} | \mathrm{X}, \lambda) = \frac{\mathrm{A_{in} B_{in}}}{\mathrm{P}(\boldsymbol{X} \mid \lambda)}. \tag{9}$$

The forward variable A_{in} is the joint probability of the past subsequence $\mathbf{X_1}, \mathbf{X_2}, \cdots, \mathbf{X_n}$ and the state $\mathbf{q_n} = \mathbf{u_i}$, i.e. $A_{in} = P(\{\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_n}\}, \mathbf{q_n} = \mathbf{u_i} | \lambda)$, given by the following recursive equation:

$$\mathrm{A_{in}} = \left(\sum_{k=1}^{M} A_{k,n-1}\, P_{ki}\right) p_i(\mathbf{x_n}) \tag{10}$$

where $A_{i,1} = \pi_i p_i(\mathbf{x_1})$. The backward variable B_{in}, which is the probability of the future subsequence $\mathbf{X_{n+1}}, \mathbf{X_{n+2}}, \cdots, \mathbf{X_N}$ given hidden state $\mathbf{q_n} = \mathbf{u_i}$, i.e. $B_{in} = P(\{\mathbf{X_{n+1}}, \mathbf{X_{n+2}}, \cdots, \mathbf{X_N}\} \mid \mathbf{q_n} = \mathbf{u_i}, \boldsymbol{\lambda})$, is calculated using the following recursive equation:

$$B_{in} = \sum_{k=1}^{M} P_{ki} p_k(\boldsymbol{x_{n+1}}) B_{k,n+1} \tag{11}$$

where B_{iN}=1.

In addition to parameters $(\mathbf{W}, \beta)$, which can be obtained in the M-step of the EM algorithm as for the standard GTM, GTM-TT modeling entails the estimation of the initial state probabilities $\{\pi_i\}$ and the state transition probabilities $\{p_{\mathrm{ij}}\}$. In order to describe the procedure for the re-estimation of this parameters, we first define $\xi_n(\mathrm{i}, \mathrm{j})$: the joint probability of hidden state $\mathbf{u_i}$ at time n and hidden state $\mathbf{u_j}$ at time $n+1$, given the data X and the model. Then, the re-estimation formulae are defined as follows:

$$\widehat{\pi_i} = R_{i1} \tag{12}$$

$$\widehat{\mathrm{P}_{\mathrm{ij}}} = \frac{\sum_{\mathrm{n}=1}^{\mathrm{N}-1} \xi_n(\mathrm{i},\mathrm{j})}{\sum_{\mathrm{n}=1}^{\mathrm{N}-1} \mathrm{R}_{\mathrm{in}}}, \tag{13}$$

The GTM is embodied with visualization capabilities that are akin to those of the SOM. The clusters of multivariate time series points can be summarily visualized in the low-dimensional latent space (in 1 or 2 dimensions) of GTM-TT by means of the posterior-mode projection (Bishop et al., 1998), defined as:

$$\mathbf{q_n^*} = \arg\max \mathrm{R}_{\mathrm{in}\ 1 \le \mathrm{i} \le \mathrm{M}} \ . \tag{14}$$

The optimum path over the space of states is defined by the state sequence $\mathrm{Q}^* = \{\mathrm{q}_1^*, \mathrm{q}_2^*, \cdots, \mathrm{q}_\mathrm{N}^*\}$. Beyond the posterior mode, the distribution of the *responsibility* over the latent space of states can also be directly visualized in full.

Appendix C. Keyword list about navigation mobile application services

Category		Keyword
Level1	Level2	
User		Drive, bike, publisher, finder, student, travelers, planner, government, marine navigator, dentist, child, organization, motorbike, consumer, coffee seeker, military, traveler, tourist, expert, army, walker
Technology	Software	Offline, direction, destination, email, current location, detection, favorite, javascript, update, text, photo, radar, drag, online, community, feedback, music, marker, language, indicator, font, path, interactive, guidance, calculation, bookmark, keyboard, measurement, landmark, sms, browser, management, audio, media, copilot, emergency, plotter, locator, status, tracker, departure, entertainment, listings, prediction, recognition, statistics, reality, shortcut, symbology, cockpit, dual, dashboard, memo, upload, translation, fax, Wikipedia, topographic, video, demo, timer, volume,

Category		Keyword
Level1	Level2	
		checkpoint, night mode, speech, symbol, reminder, graphic indicator, protector, safari, realtime, photo library, visual warning, brightness, tv, recorder, amusement, instructor, summary, lane assistance, emails, architecture, enroute, screenshot, clipboard, privacy, playback, toolbar
	Hardware	Gps, internet connection, camera, wifi, altimeter, battery, rotation, speedometer, fingertips, storage, vibration, velocity, receiver, resolution, accelerometer, car charger, magnetometer, acceleration, odometer, sensor, projection, restart, daylight, aux, windscreen, fullscreen, toggle, sunlight, flashlight
Information	Location	Location, street, area, restaurant, city, campus, country, road, lake, subway, college, metro, university, museum, airport, building, region, aircraft, hall, hotel, speedcam, food, office, dorms, entry, coffee, movie, airplane, mall, company, attraction, medicine, aviation, party, river, furniture, outlet, theater, eat,

Category		Keyword
Level1	Level2	
		rental car, highway, facility, beauty, ocean, town, safety camera, hamburg, snorkeling, mountain, stairway, starbucks, pharmacy, gallery, buddy, valley, casino, county, pizza, charity, grocery, atms, hospital, seafood, winery, festival, electronics, comedy, education, karaoke, motel, takeout, tesco, art gallery, laundry, dinner, delivery, neighborhood, steakhouse, suburb, theatre, studio, Disney, harbor, factory, motorway, car rental, fitness, cinema, sushi, drink, Hilton, bakery, bookstore
	Object	Width, car, height, altitude, coordinates, latitude, longitude, elevation, depth, instruction, situation, grid, subscription, news, transportation, vessel, engine, hotspot, coordinate, bart, aero, airway, football, reputation, magazine, hockey, yoga, timetables, opera, tennis, postcode, public transportation, seaplane, overview, sunset, sunrise, soccer, promotion, dvd, incident, temperature

Appendix D. Keyword list about camera technology-based mobile application service

No.	Keyword	No.	Keyword
1	text	14	adjustment
2	location	15	resize
3	timer	16	fax
4	exposure	17	scan
5	augmented reality	18	qr
6	draw	19	health
7	photography	20	postcard
8	movie	21	saturation
9	video	22	ocr
10	portrait	23	document
11	direction	24	fun
12	brightness	25	flash

초 록

정보시각화의 주요 목적은 어떠한 정보를 시각적 형상으로 표현함으로써 그 정보를 묘사하고 탐색하는 프로세스를 구축하는 것으로써 방대한 양의 정보를 분석할 때 정보시각화의 활용은 여러 가지 장점을 가지고 있다. 즉, 시각화는 인간의 인지능력을 증폭시키고 특정한 활동을 수행하기 위해서 필요한 복잡한 인지과정을 줄여준다. 또한, 큰 그림을 보여줌으로써 통찰력을 제공하기도 한다. 따라서 본 학위논문은 기술경영의 변화에 따라서 계속적으로 그 중요성이 더해지고 있는 기술 및 서비스 영역에서의 지금까지 개발되지 않은 공백을 발견하고 그 트렌드를 분석하기 위해서 정보시각화를 이용한 기술지도를 개발하고 활용한다. 이는 군의 무기체계 개발 및 획득에 필요한 기술개발에 활용할 수 있어서 국방력 강화에 크게 이바지할 수 있을 것으로 판단된다. 본 학위논문은 3개의 연구로 구성된다. 3개의 연구 안에서 1) Generative topographic mapping (GTM) 기반의 특허 공백지도를 이용해서 기술 공백을 발견하고 2) GTM 기반의 서비스 공백지도를 이용해서 서비스 공백을 발견하며 3) Generative topographic mapping through time (GTM-TT) 기반의 서비스 트렌드 지도를 활용해서 서비스 트렌드를 분석하는 접근법을 제시한다. 방대한 양의 문자 데이터로부터 중요한 정보를 추출하고 분석하기 위해서 텍스트 마이닝 기법을 바탕으로 벡터 공간 모형을 활용하여 비구조화된 문서를 구조화된 자료로 변환한다. GTM은 다차원의 데이터 공간을 저차원의 잠재공간으로 그리고 그 역방향으로 사상시킬 수 있는 확률적 모델이며 베이지안 이론에 기초한 확률적 방법을 활용한다는 측면에서 self organizing map (SOM)의 대용 모델이라고 할 수 있다. GTM-TT는 시간 기반 군집화와 시각화를

동시에 수행함으로써 다변량 시계열자료의 탐색적 분석을 위한 GTM의 확장 모델이다.

첫 번째 연구는 GTM 기반의 특허 공백지도를 이용하여 특허공백을 파악하고 이를 통해서 새로운 기술 기회들을 발견하는 문제를 다룬다. 특허지도는 잠재되어 있는 기술적 정보를 얻기 위한 유용한 도구로써 오랫동안 인식되어 왔다. 다른 영역 중에서도 새로운 기술의 탐색되지 않은 영역으로 정의되는 특허공백을 발견하고자 하는 연구영역이 주목을 받아왔다. 그러나 이전의 연구들에서는 특허지도에서 특허공백을 발견함에서 있어서 특허공백을 주관적으로 그리고 수동적으로 파악해야 하는 한계가 있어왔다. 따라서, 본 연구에서는 특허공백을 자동적으로 그리고 객관적으로 발견하기 위해서 GTM 기반 특허 공백지도라는 기술지도를 제안한다. GTM은 다차원의 데이터 공간을 저차원의 잠재공간으로 사상하고 그 역사상이 가능하므로 특허공백을 자동으로 발견하고 해석하는데 기여한다. 제안하는 접근법은 크게 3가지 단계로 구성되어 있다. 첫 째, 텍스트 마이닝을 이용해서 특허문서들을 구조화된 데이터인 키워드 벡터로 변환한다. 둘 째, GTM을 적용하여 특허지도를 만들고 지도에서 비어있는 영역으로 표현되는 특허공백을 발견한다. 마지막으로, 특허지도를 키워드 벡터로 역사상하여 특허공백을 새로운 기술 기회로 해석한다. 사례연구는 반도체 공정에서 필요한 리소그라피 (lithography) 기술 관련된 특허를 대상으로 실시한다. 본 연구는 특허공백을 발견하여 새로운 기술 기회를 발견하기 위한 시간과 노력을 절약할 수 있을 뿐만 아니라 객관성과 신뢰성을 증진시킬 수 있을 것으로 기대된다.

두 번째 연구는 기술 영역을 다루었던 첫 번째 연구와는 달리 GTM 기반 서비스 공백지도를 이용하여 서비스 공백을 도출함으로써 새로운 서비스 기회를 발견하고자 하는 문제를 다룬다. 새로운 서비스 기회에 대한 전략적 기술적 중요성에도 불구하고 새로운 서비스 기회의 지능적 탐색과 체계적 발견에 관한

연구가 상대적으로 부족하였다. 본 연구에서는 새로운 서비스 기회를 발견하기 위한 GTM 기반 서비스 지도를 개발하고 활용하는 접근법을 제안한다. 스마트폰 사용자들에게 모바일 애플리케이션 서비스에 접근할 수 있도록 하는 웹서비스를 제공하는 애플의 앱스토어 (AppStore) 로부터의 모바일 애플리케이션 서비스를 활용하여 사례연구를 실시한다.

세 번째 연구는 GTM-TT 기반의 서비스 트렌드 지도를 개발하고 이를 이용하여 서비스의 트렌드를 파악하고자 한다. 최근, 서비스의 폭발적인 증가로 인해 기업들은 직관적이고 객관적으로 서비스의 패턴과 트렌드를 분석해야 하는 문제에 직면해 있다. 이러한 상황에서 서비스 트렌드 지도라는 것이 활용될 수 있다. 일반적으로 지도는 사람들이 방대한 양의 정보를 시각화할 수 있는 시각화 도구로서의 잠재성 때문에 상당한 주목을 받아오고 있다. 특히, GTM-TT는 시간 기반 군집화와 변화 경로를 제공함으로써 동태적 분석에 적합한 모델이다. 따라서 본 연구에서는 서비스의 트렌드를 분석하기 위해서 서비스 군집 지도와 서비스 경로 지도로 구성된 GTM-TT 기반 서비스 트렌드 지도를 제안한다. 본 연구는 크게 데이터베이스 구축, 데이터 전처리, GTM-TT 서비스 트렌드 지도의 개발, 해석의 4가지 단계로 구성되어 있으며 다른 서비스 영역에서도 동태적 서비스 트렌드를 파악하는데 도움을 줄 수 있을 것으로 기대한다.

주요어: 정보시각화, 텍스트 마이닝, generative topographic mapping (GTM), generative topographic mapping through time (GTM-TT)

학 번: 2009-30757

감사의 글

논문을 마치고 '감사의 글'을 접하고 보니 지난 대학원 생활이 주마등처럼 뇌리를 스쳐지나 갑니다. 대한민국 육군 장교와 대학원생 그리고 한 가정의 가장으로서의 역할을 다하기 위해서 끊임없이 노력했던 제 모습과 그런 저를 곁에서 지켜보며 격려해주신 많은 분들과 함께 뜻 깊은 대학원 생활을 마무리 하려고 하니 만감이 교차합니다. 부족한 제가 이렇게 학위논문을 세상에 떳떳하게 내놓을 수 있게 된 것은 많은 고마운 분들 덕분이며 그 모든 분들에게 감사의 마음을 전하고자 합니다.

먼저 지도교수님이신 박용태 교수님께 감사드립니다. 패기만을 가지고 시작한 박사과정이었지만 교수님의 세심한 배려과 관심, 따뜻한 사랑과 학문적 가르침을 통해서 이렇게 성장할 수 있었습니다. 그리고 교수님께서 보여주신 참된 스승의 모습은 제 가슴에 깊이 새겨졌습니다. 또한 부족한 논문을 지도해주시고 심사해주신 윤명환 교수님, 박종헌 교수님, 박우진 교수님, 윤병운 교수님께 깊은 감사를 드립니다.

기술경영연구실 가족들에게도 감사의 마음을 전합니다. 항상 든든한 버팀목이 되어주신 박광만, 김문수, 설현주, 강인태, 윤병운, 신준석, 최창우, 이성주 박사님께 감사를 드립니다. 또한 3년 반이라는 시간 동안 동고동락했던 정환형, 학연, 승겸, 소라, 영정, 창용, 상훈, 용윤, 우리, 양래, 효진, 대국, 보미, 란, 지은, 유진, 성수, 은지, 지영, 우석, 현정에게도 고마움을 전합니다. 함께 연구실 생활은 하지 않았지만 따뜻하게 대해주신 은철선배, 경도형, 낙환형도 감사합니다. 연구실

가족들과의 인연은 제가 대학원에서 얻은 가장 소중한 자산입니다.

박사과정을 무사히 마칠 수 있었던 것은 힘들 때마다 가까운 곳에서 용기를 북돋아 준 친구들과 동료들이 있었기 때문입니다. 초등학교부터 지금까지 끈끈한 우정을 과시한 항제와 태준이, 육군사관학교 58기 동기생 및 선후배들, 함께 위탁교육을 받았던 육.해.공군 선후배 및 동기들, 협동과정의 효정누나, 연구실은 다르지만 함께 테니스도 치고 술잔도 기울였던 홍범에게 감사의 마음을 전합니다. 마지막으로, 멀리에서도 항상 응원을 아끼지 않으신 3사관학교의 동료 및 선후배 교수님들, 특히, 무기시스템학과 교수님들께 감사드립니다.

그리고 무엇보다도 무조건적인 사랑과 헌신으로 키워주신 아버지와 하늘나라에서 지켜보며 기뻐하실 어머니, 정말 감사하고 사랑합니다. 부족한 사위를 항상 아껴주시고 걱정해주신 장인어른, 장모님 그리고 처남께도 깊은 감사를 드립니다. 사랑하는 형과 누나, 매형, 그리고 예쁜 우리 조카들, 수민, 수경에게 늘 고마운 마음을 간직하고 있습니다. 사촌형, 누나들과 동생들, 외할머니, 이모, 고모할머니, 고모할아버지, 영천 큰아버지, 큰어머니, 반야월 큰어머니, 삼촌께도 감사를 드립니다. 마지막으로, 힘든 세월을 묵묵히 저의 곁을 지켜준 사랑하는 아내 정은과 소중한 세 아들 준영, 우영, 호영에게 이 논문을 바칩니다.

2012년 8월

관악에서

손창호 올림