Expert Systems with Applications 39 (2012) 2489–2500

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Development of a GTM-based patent map for identifying patent vacuums

Changho Son, Yongyoon Suh, Jeonghwan Jeon, Yongtae Park *

Department of Industrial Engineering, School of Engineering, Seoul National University, San 56-1, Shilim-Dong, Kwanak-Gu, Seoul 151-742, Republic of Korea

ARTICLE INFO

Keywords:
GTM
Patent map
Patent vacuum
Keyword vector

ABSTRACT

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 paper proposes a generative topographic mapping (GTM)-based patent 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 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.

1. Introduction

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; Ernst, 2003; Griliches, 1990; Jaffe, Trajtenberg, & Forgarty, 2000; Li, Wang, & Hong, 2009), the patent map is a guaranteed useful proxy measure for technological power (Park, Yoon, & Lee, 2005), has been employed as the representative tool used to grasp diverse features of individual patents and identify complex relationships among patents (Yoon, Yoon, & Park, 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, Yoon, & Park, 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.

* Corresponding author. Tel.: +82 2 880 8358; fax: +82 2 889 8560.
E-mail addresses: c13981@snu.ac.kr (C. Son), yue2000@snu.ac.kr (Y. Suh), hwan63@snu.ac.kr (J. Jeon), parkyt@cybernet.snu.ac.kr (Y. Park).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.08.101

This paper proposes a generative topographic mapping (GTM)-based patent 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 (Bishop, Svensén, & Willams, 1998), the contributions to the detection and interpretation of patent vacuums are twofold. Firstly, in regards to detection, the GTM-based patent 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 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 of patent vacuums and subjective interpretation of patent vacuums.

This paper is structured as follows. Section 2 addresses the underlying methodology for the proposed approach: patent map and GTM. Section 3 focuses on the overall research framework and detailed processes for GTM-based patent map used to detect and interpret patent vacuums. The case study with lithography technology-related patents is provided in Section 4, finally followed by the discussion and conclusion.

2. Literature review

2.1. Patent analysis

A variety of sectors have extensively employed patent analysis, including entire nations, industries, firms, and technological fields. Patents possess useful information and effectively act as a public database, which is documented and organized in standardized formats (Wartburg, Teichert, & Rost, 2005). Among others advantages, patents present an ample source of technical and market information such as technical features, ownership, and commercial worth (Kuznets, 1962; Park et al., 2005; Soo, Lin, Yang, Lin, & Cheng, 2006). Reports indicate that patents demonstrate strong correlations to a firms' success, and include about 80% of the world's technology knowledge (Ernst, 2001; Lerner, 1994). The purpose of patent analysis is diverse in terms of technical and economic decision making: technology forecasting (Morris, DeYong, Wu, Salman, & Yemenu, 2002; Yoon & Park, 2007), technology evolution analysis (Choi & Park, 2009), or technology trend analysis (Basberg, 1987). In fact, a technology strategy in firms such as a technology acquisition, a technology transfer, and even a merger and acquisition can be formulated with methodological patent analysis above (Narin, Noma, & Perry, 1987).

In particular, patent analysis is comprised of both structured data analysis and unstructured data analysis. The value of patent analysis has rapidly grown due to the simultaneous management of both structured and unstructured. The methods and techniques for patent analysis are applied in correspondence to the characteristics of structured and unstructured data. Firstly, structured data is comprised of a consistent and standardized format such as a patent number, a filing date, an issued date, a cited patent, inventors, or assignees (Lee, Yoon, Lee, & Park, 2009; Verbeek et al., 2002). Typically, structured data has been analyzed by means of simple, descriptive statistics such as a graph, a chart, a bar, and a table. A bibliometric technique is used for investigating dynamic trends in technology development, and statistical methods make structured data more informative and constructive (Yoon et al., 2002). Citation network analysis is increasing in popularity as an advanced structured data analysis for monitoring technological developments such as organic technology evolution (Choi & Park, 2009; Meyer, 2000) and interdisciplinary technology fusion (No & Park, 2010). Secondly, unstructured data indicates texts or descriptions of patent documents; and, descriptions found within unstructured data contain the main ideas for technological development or innovation. However, since it is difficult to systematically extract meaningful implications from natural languages, such as texts or descriptions, unstructured data must inevitably transform into structured data. A text mining technique is a widely used technique that logically and automatically derives keywords from collections of unstructured data (Kostoff, Toothman, Eberhart, & Humenik, 2001). In relation to patent analysis, the text mining technique encompasses a significant role in data-preprocessing and information-extracting. The extracted keywords from patents usually include core technology, product, component, and methods. Thus, this technique is extensively utilized for identifying patent keywords and exploring the combinations among keywords for new technologies (Yoon & Park, 2005).

2.2. Patent map

The patent map embraces a multitude of visual concepts and descriptions based on the relationships among patents, such as charts, graphs, bars, and tables (Chen, 2009; Liu, 2003). The patent map is an effective visualization technique for formulating strategies because it provides practical and intuitive information (Kim, Suh, & Park, 2008). Therefore, developing a patent map is imperative for graphically providing invaluable information, as well as exploring technological opportunities through patent documents. Since text mining is applied to patent maps when evaluating unstructured data, patent maps have effectively communicated potentially beneficial knowledge and explicit information from patents (Larkey, 1999; Tseng, Lin, & Lin, 2007; Tseng, Wang, Lin, Lin, & Juang, 2007). Patent maps are categorized according to specific purposes: the technical patent map, the management patent map, and the claim patent map (Yoon et al., 2002). The technical map is used to understand core technology, as well as identify potential technology. Management patent maps trace dynamic trends of specific technology. Claim patent maps are useful for monitoring patent conflicts. In particular, technical patent maps are especially useful in discovering patent vacuums through unexplored patent data on the map. Patent vacuums help practitioners formulate future plans through identifying important potential technology. Thus, the principle component analysis (PCA) and the self-organizing map (SOM) are representative techniques that locate patents in a patent map.

PCA is a form of statistical analysis used for dimensionality reduction, which is accomplished by converting multi-variables into a few linear combinations (Johnson & Wichern, 1988). PCA provides the opportunity to understand latent dimensions due to the fact that significant principle components (PCs), which cause "variance" in data, can be extracted. Data can be represented by significant dimensions. However, it is hard to interpret the meaning of new dimensions since one PC contains too much dimensional data.

The SOM is an artificial intelligence technique utilized for visualizing multi-dimensional data as two-dimensional space neurons (Kohonen, 1998). The SOM employs an algorithm in which interrelated vectors are heuristically grouped as a neighborhood and

regarded in terms of artificial learning parameters. There are two major advantages of the SOM. Firstly, SOM is able to identify similar technologies since each patent is located on a single, discrete node. Secondly, the ability of SOM to visualize data as set of neurons is a powerful, presenting similarities and differences in data by color contrasts in patent maps (Kohonen, 1995). Artificial learning processes, which update neighboring nodes and the weights, create simplified images of the observable real world (Kohonen, 1982). In spite of the two strengths that SOM possesses, there is a lack of theoretical proof as to why the patent is mapped on a specific node due to the unsupervised machine-learning process (Bishop et al., 1998). Since the SOM training algorithm optimizes an objective function by a heuristic approach, data over-fitting also remains a problem (Svensén, 1998)

As for the patent map, two main methods were employed to identify patent vacuums. Lee, Yoon, and Park (2009) and Yoon et al. (2002) proposed the PCA-based patent map and the SOM-based patent map, respectively. Each map has distinct characteristics and applications with respect to the strengths of the PCA and the SOM. More specifically, the patent map was developed in order to investigate patent vacuums as technological opportunities. Vacuums in the patent map are represented as sparse areas; potential new technology can be explored in each vacuum. However, during this the PCA and the SOM processes, patent vacuums are detected and interpreted only through subjective judgment in which patents surrounding the patent vacuums are the main interest for investigation due to limited information about vacuums. Because expert knowledge is an unavoidable resource for richer descriptions of the technology in patent vacuums, a heavy dependence on subjective decision-making results. Fundamental limitations still exist in the aforementioned mapping techniques, as well as other techniques, in regards to demonstrating characteristics of vacuums without the help of experts. Therefore, it is vital to develop the best system for identifying and exploring patent vacuums automatically and objectively.

2.3. Generative topographic mapping

2.3.1. 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 et al. (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 1.

2.3.2. 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 Fig. 1. The x indicates a reduced data vector in the latent space; R^L and t represent an observed data vector in the data space, R^D.

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

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

y: transformation function, x: latent variables, t: data variables, D: dimension of t, $y(x,W)$: transformed x into data set and 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

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

However, it is difficult to deduce $p(t)$ because it is continuous distribution. To overcome this issue, the delta function is applied to discrete $p(t|x)$. 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. Fig. 2 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)

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

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

Table 1
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

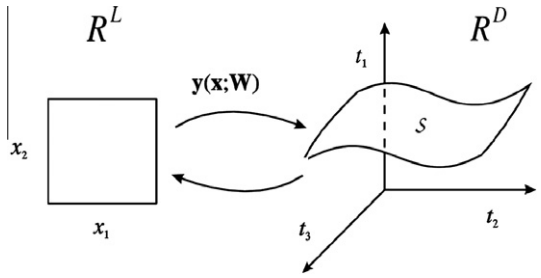


Fig. 1. Basic concept of GTM.

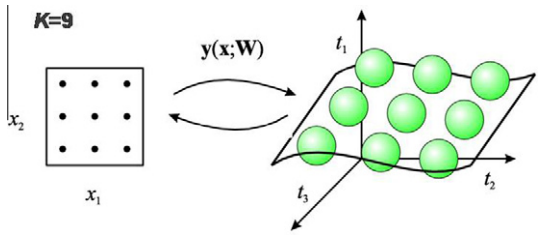


Fig. 2. Mapping regular grids into the data space.

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)

$$p(x_k|t_n, W, \beta) = \frac{p(t_n|x_k, W, \beta)p(x_k)}{\sum_{k'} p(t_n|x_{k'}, W, \beta)p(x_{k'})} \tag{4}$$

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

$$y(x, W) = \Phi(x)W \tag{5}$$

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

3. Identification of patent vacuums

3.1. Overall research framework

Fig. 3 depicts the overall research framework, which consists of several stages. Firstly, patent documents related to technology under consideration are collected from the US 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.

3.2. Detailed processes

3.2.1. 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. 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 2. 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.

However, since the objective of this paper is to identify the patent vacuums and not to investigate patent trends, binary keyword vectors were employed instead of the frequency of keyword

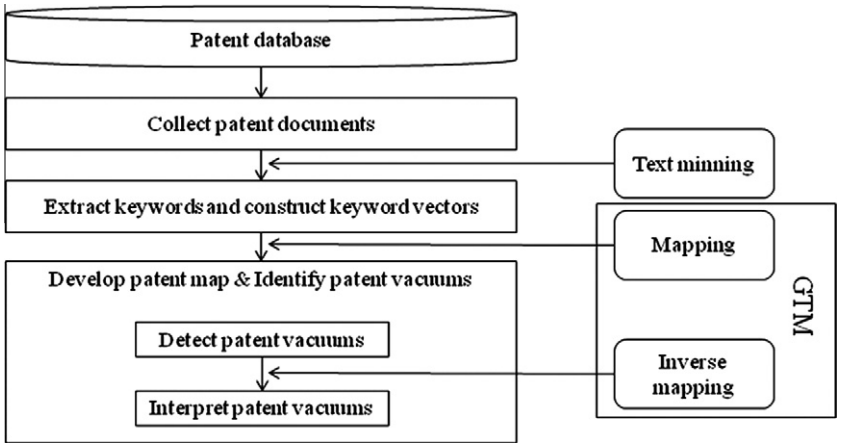


Fig. 3. Overall research framework.

Table 2
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

occurrence vector. For instance, if patent 1 has keyword 1 and keyword n, two fields are filled with "1", respectively.

3.2.2. Development of GTM-based patent map

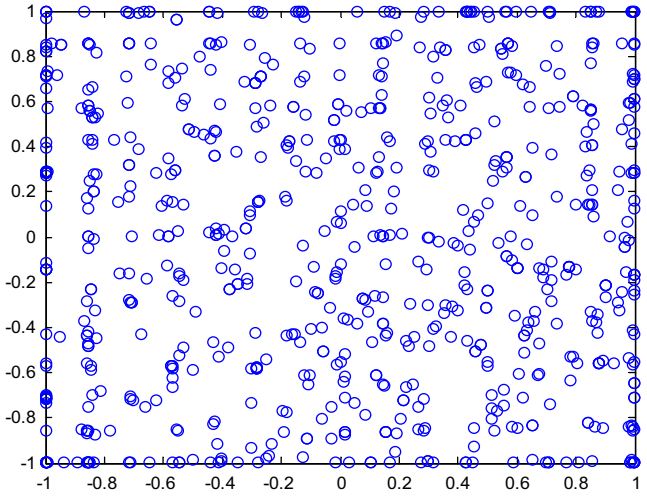
Subsequently, the GTM-based patent map is developed from the constructed keyword vectors. If fifty keywords are extracted, each keyword vector has 50 dimensions. This presents difficulties in both visualizations and interpretations. Therefore, it is necessary to visualize the vectors in two-dimensional 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 map is constructed, as illustrated in Fig. 4. Fig. 4(a) shows the posterior-mean projection of the data in the latent space and Fig. 4(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 '○' in Fig. 4(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.

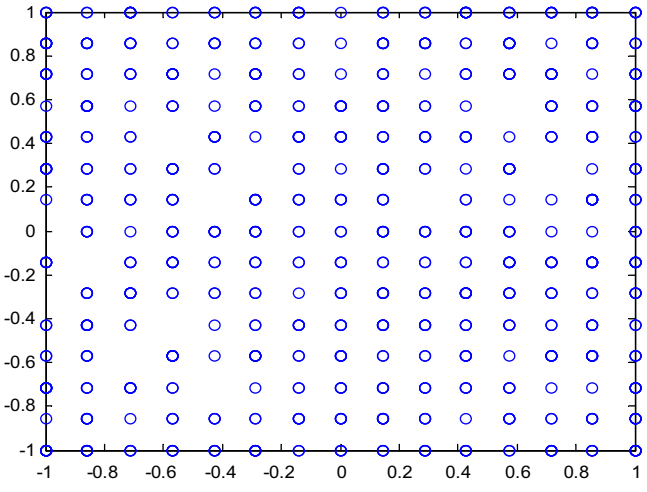
3.2.3. Detection of patent vacuums

In the GTM-based patent map, patent vacuums are identified as the blank areas in the map. As shown in Fig. 5, the blank grid, which is represented by an X in red,[1] are identified as patent vacuums. Since the GTM-based patent map mainly consists of grids and each patent is located at each grid, the blank grid is intuitively

[1] For interpretation of color in Fig. 5, the reader is referred to the web version of this article.



(a) The posterior-mean projection of the data



(a) The posterior-mode projection of the data

Fig. 4. An example of the GTM-based patent map.

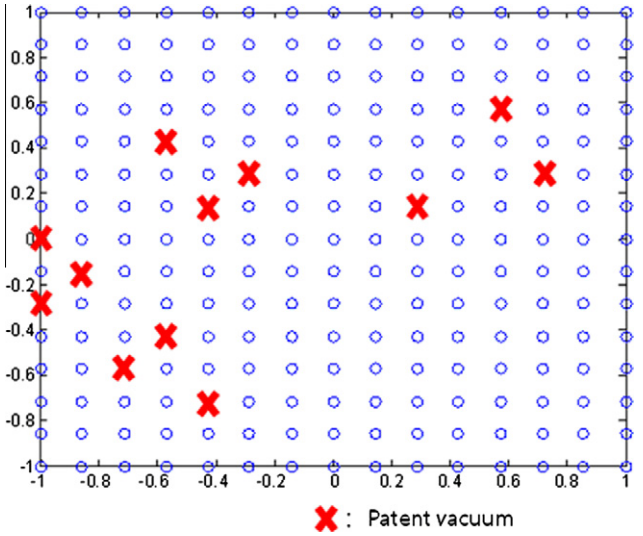
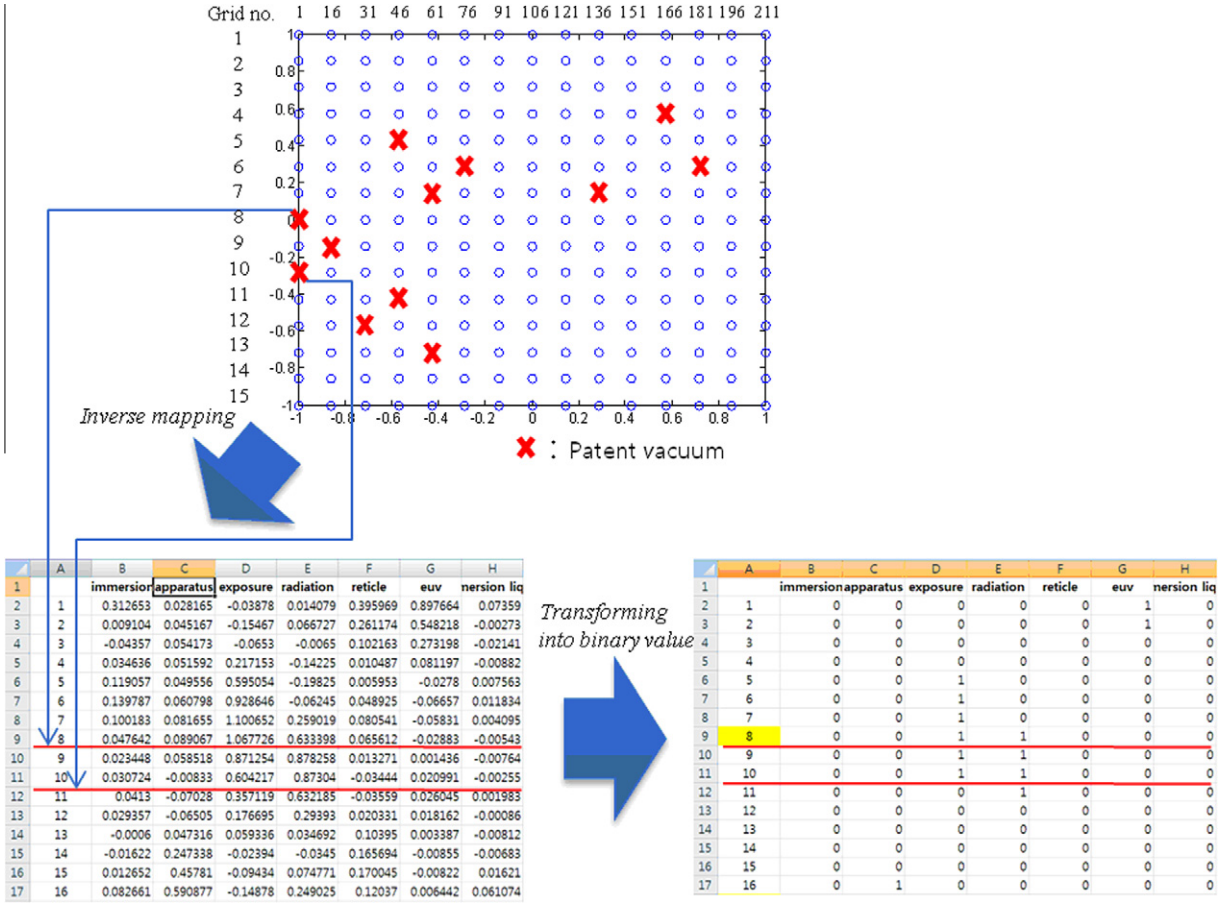


Fig. 5. An example of patent vacuums.



	A	B	C	D	E	F	G	H
1		immersion	apparatus	exposure	radiation	reticle	euv	nersion liq
2	1	0.312653	0.028165	-0.03878	0.014079	0.395969	0.897664	0.07359
3	2	0.009104	0.045167	-0.15467	0.066727	0.261174	0.548218	-0.00273
4	3	-0.04357	0.054173	-0.0653	-0.0065	0.102163	0.273198	-0.02141
5	4	0.034636	0.051592	0.217153	-0.14225	0.010487	0.081197	-0.00882
6	5	0.119057	0.049556	0.595054	-0.19825	0.005953	-0.0278	0.007563
7	6	0.139787	0.060798	0.928646	-0.06245	0.048925	-0.06657	0.011834
8	7	0.100183	0.081655	1.100652	0.259019	0.080541	-0.05831	0.004095
9	8	0.047642	0.089067	1.067726	0.633398	0.065612	-0.02883	-0.00543
10	9	0.023448	0.058518	0.871254	0.878258	0.013271	0.001436	-0.00764
11	10	0.030724	-0.00833	0.604217	0.87304	-0.03444	0.020991	-0.00255
12	11	0.0413	-0.07028	0.357119	0.632185	-0.03559	0.026045	0.001983
13	12	0.029357	-0.06505	0.176695	0.29393	0.020331	0.018162	-0.00086
14	13	-0.0006	0.047316	0.059336	0.034692	0.10395	0.003387	-0.00812
15	14	-0.01622	0.247338	-0.02394	-0.0345	0.165694	-0.00855	-0.00683
16	15	0.012652	0.45781	-0.09434	0.074771	0.170045	-0.00822	0.01621
17	16	0.082661	0.590877	-0.14878	0.249025	0.12037	0.006442	0.061074

	A	B	C	D	E	F	G	H
1		immersion	apparatus	exposure	radiation	reticle	euv	nersion liq
2	1	0	0	0	0	0	1	0
3	2	0	0	0	0	0	1	0
4	3	0	0	0	0	0	0	0
5	4	0	0	0	0	0	0	0
6	5	0	0	1	0	0	0	0
7	6	0	0	1	0	0	0	0
8	7	0	0	1	0	0	0	0
9	8	0	0	1	1	0	0	0
10	9	0	0	1	1	0	0	0
11	10	0	0	1	1	0	0	0
12	11	0	0	0	1	0	0	0
13	12	0	0	0	0	0	0	0
14	13	0	0	0	0	0	0	0
15	14	0	0	0	0	0	0	0
16	15	0	0	0	0	0	0	0
17	16	0	1	0	0	0	0	0

Fig. 6. An example of inverse mapping.

identified as a vacuum. Thus, manual work conducted by researchers is unnecessary for identifying patent vacuums in the GTM-based patent map.

3.2.4. 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 Fig. 6. 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 Fig. 6. 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.

4. 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 (Harriott, 2001; Stulen & Sweeney, 1999). 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 & 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 (Levinson, 2001; Smith, 1998).

4.1. Data collection

As mentioned above, lithography technology patents are the underlying source for data presented in this paper. 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.

Table 3
Extracted keywords.

No.	Keywords	No.	Keywords	No.	Keywords	No.	Keywords
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

4.2. Data preprocessing

Since most text-mining algorithms use keywords for expressing the context of the document (Yoon & Park, 2004), this paper 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.

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 Fig. 7.

4.3. Development of GTM-based patent map

After data preprocessing, GTM was employed to develop the GTM-based patent map for identifying patent vacuums. Prior to developing the GTM-based patent 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 map is developed using MATLAB R2008a with GTM toolbox developed by Svensén (1998) as shown in Fig. 8.

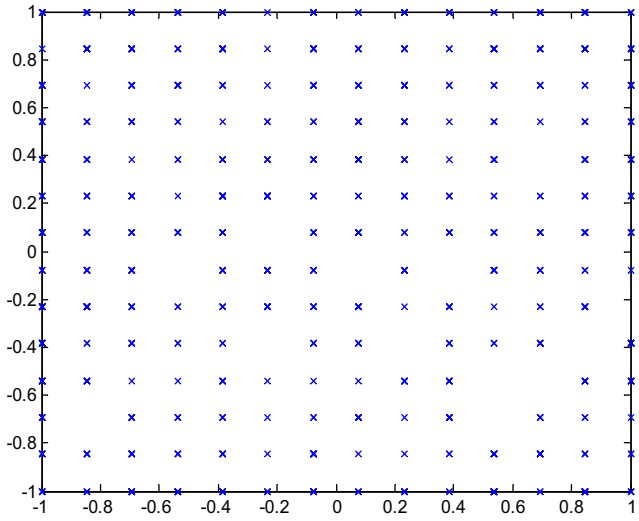


Fig. 8. GTM-based patent map (the posterior-mode projection).

4.4. 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. Fig. 9 shows the patent vacuum identified from the GTM-based patent map. A total of 13 patent vacuums of 169 latent points were discovered through posterior-mode projection.

4.5. Interpretation of patent vacuums

Inverse mapping was conducted in order to interpret the meaning of the identified patent vacuums. Each vacuum in Fig. 9 was transformed into the keyword vector as a means to represent the

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)

Fig. 7. Keyword vector construction.

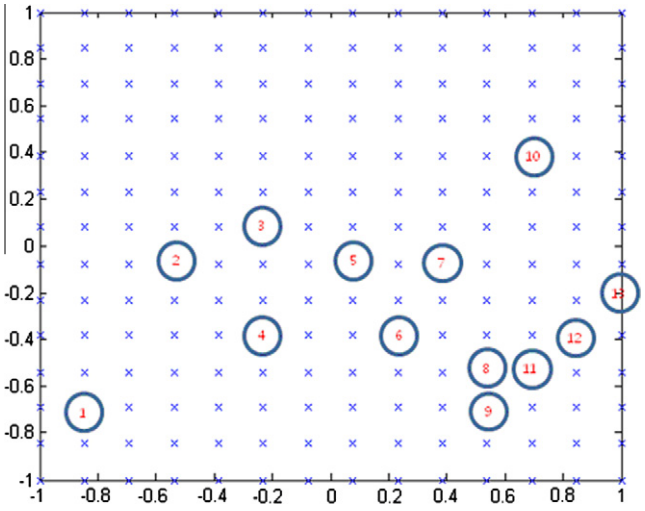


Fig. 9. Patent vacuums identified from GTM-based patent map.

characteristics of each vacuum. The results of the inverse mapping are represented in Fig. 10.

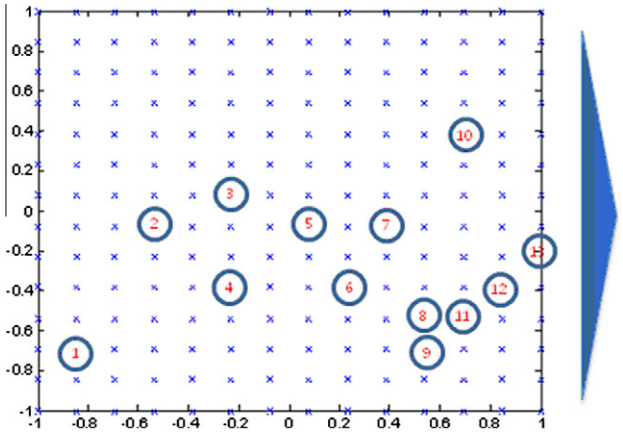
Since the original keyword vector is composed of binary values, the result of inverse mapping should also be composed of binary values. Therefore, each field of keyword vector obtained from the inverse mapping was filled with a 0 or 1 value according to the threshold, as shown in Table 4. The threshold 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 map can be interpreted in the level of keywords, which is the form of the original dataset.

Table 5 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.

5. Discussion

The GTM-based patent map needs to be validated for practical use in the field, and to show that this outperforms 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. Fig. 11 shows PCA-based patent map, SOM-based patent map, and GTM-based patent map, respectively.



	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

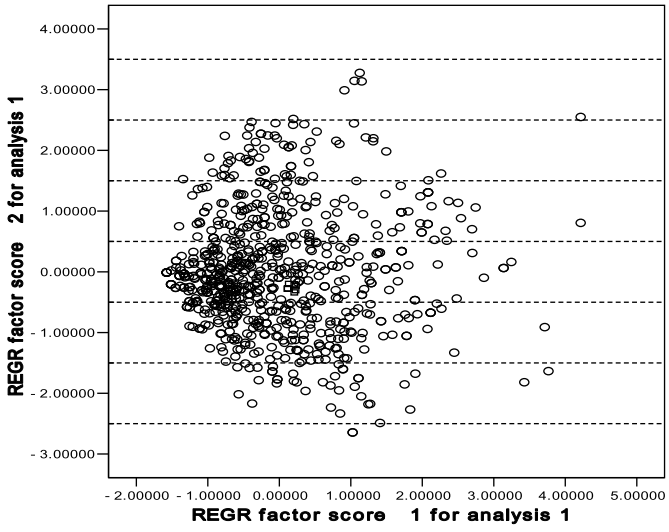
Fig. 10. The result of inverse mapping.

Table 4
Keyword vector of patent vacuums.

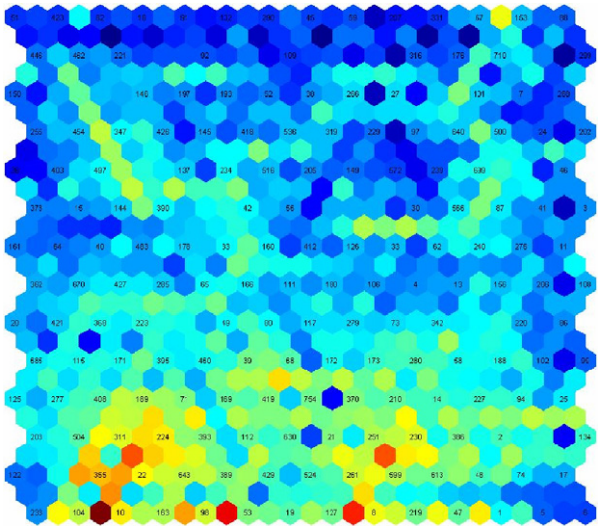
Vacuum no.	Keywords																																					Sum
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	
1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
2	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0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

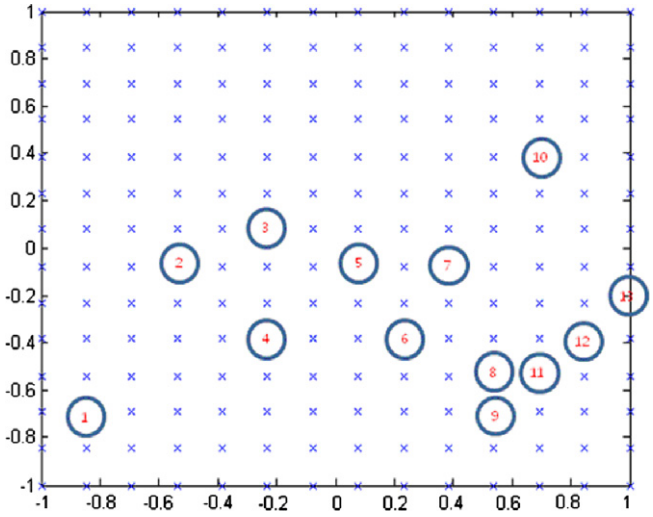
Patent vacuum no.	Keywords
1	Apparatus, exposure, lens
2	Apparatus, exposure, alignment
3	Apparatus, exposure, radiation, reticle, wavelength, alignment
4	Apparatus, exposure, radiation, reticle, laser, axis, calibration,
5	Apparatus, exposure, reticle, axis, wavelength, lens
6	Apparatus, exposure, reticle, 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



(a) PCA-based patent map



(b) SOM-based patent map



(c) GTM-based patent map

Fig. 11. Three distinctive patent maps.

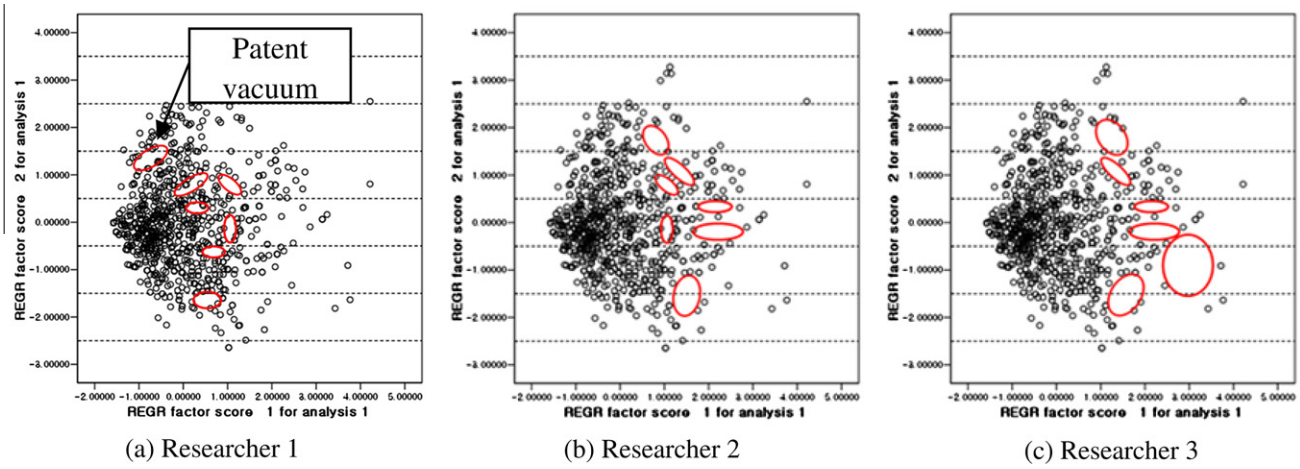


Fig. 12. Patent vacuums depending on researchers in PCA-based patent map.

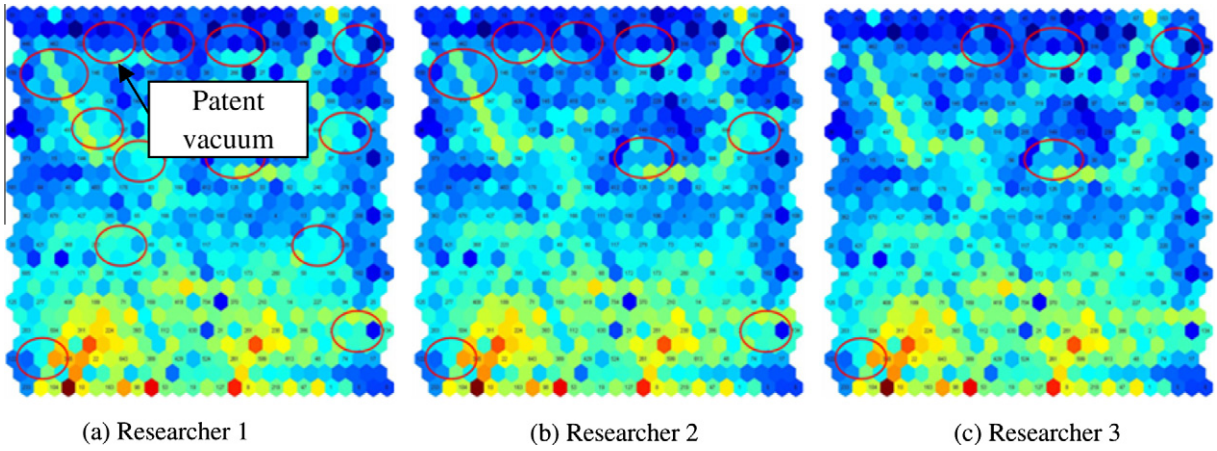


Fig. 13. Patent vacuums depending on researchers in SOM-based patent map.

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 Figs. 12 and 13. Sparse areas in the PCA-base patent map in Fig. 11(a) are considered as patent vacuums. However, patent vacuums represented ellipses might be changed according to researchers' judgments as shown in Fig. 12 since the definition of sparseness varies depending on the researchers. In Fig. 11(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 Fig. 13. However, patent vacuums in the GTM-based patent map are automatically detected using grid-based visualization since a blank grid is considered as a patent vacuum as shown in Fig. 11(c).

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 Figs. 12(a) and 13(a) are shown in Table 6. 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 6
List of surrounding patents of target patent vacuum.

PCA-based patent map			SOM-based patent map		
Surrounding patents of target patent vacuum					
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 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 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 map is more appropriate for identifying patent vacuums among lots of patents since it can automatically and objectively detect and interpret patent vacuums.

6. Conclusions

Although identifying patent vacuums has been considered an important issue in regards to exploring new technology, relatively little research has been devoted to the visual exploration and identification of patent vacuums. More importantly, existing patent maps used for identifying patent vacuums bear the two significant limitations: the subjective detection and subjective interpretation of patent vacuums. In response to these limitations, this paper proposed a GTM-based patent map to automatically identify patent vacuums, ultimately compensating for the aforementioned limitations.

The contributions of the suggested GTM-based patent map are clear: Firstly, the automatic and objective detection of patent vacuums compared to the previous techniques is a result of the GTM-based patent 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.

Specifically, we demonstrated how to develop the GTM-based patent map with lithography technology-related patents as well as how to automatically identify patent vacuums from the GTM-based patent map. Thus, researchers, engineers and managers interested in new technology development save time and energy when uncovering new technology opportunities as well as acquire objective results.

Despite the comprehensive and objective aspects of the GTM approach, this technique still possesses its limitations. Firstly, a necessity in elaborating the keyword extraction process is vital, since keyword extraction plays a critical role in determining the value of patent vacuums. Although the text mining tool and expert judgment were employed to extract the keywords, covering both quantitative qualitative perspectives of keyword extraction, other systematic methodologies should supplement these techniques in order to validate the extracted keywords. Secondly, GTM is very sensitive to parameter settings potentially resulting in inappropriate patent vacuums if the parameters are set incorrectly. In this paper, sensitive analysis was used to determine the parameters. Even the sensitive analysis can be an alternative to overcome the sensitivity. A systematic guideline for defining parameters is potentially another subject matter requiring further study that would complement the progress made with GTM-based patent map.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2009-0085757).

References

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

Basberg, B. L. (1987). Patents and the measurement of technological change: A survey of literature. *Research Policy, 16*(2/4), 131–141.

Bishop, C., Svensén, M., & Willams, C. K. I. (1998). GTM: The generative topographic mapping. *Neural Computation, 10*(1), 215–234.

Chen, R. (2009). Design patent map visualization display. *Expert Systems with Applications, 36*(10), 12362–12374.

Choi, C., & Park, Y. (2009). Monitoring the organic structure of technology based on the patent development paths. *Technological Forecasting and Social Change, 76*(7), 769–786.

Ernst, H. (2001). Patent applications and subsequent changes of performance: Evidence from time-series cross-section analyses on the firm level. *Research Policy, 30*(1), 143–157.

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

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

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.

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

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

Jaffe, A. B., Trajtenberg, M., & Forgarty, M. S. (2000). Knowledge spillovers and patents citations: Evidence from a survey from inventors. *American Economic Review, 90*(2), 215–218.

Janakiram, M., & Goernitz, S. (2005). Real-time lithography registration, exposure, and focus control – A framework for success. *IEEE Transactions on Semiconductor Manufacturing, 18*(4), 534–538.

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

Kim, Y., Suh, J., & Park, S. (2008). Visualization of patent analysis for emerging technology. *Expert Systems with Applications, 34*(3), 1804–1812.

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

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

Kohonen, T. (1998). *Visual explorations in finance with self-organizing maps*. Berlin: Springer.

Kostoff, R., Toothman, D., Eberhart, H., & Humenik, J. (2001). Text mining using database tomography and bibliometrics: A review. *Technological Forecasting and Social Change, 68*(3), 223–252.

Kuznets, S. (1962). Innovative activity: Problems of definition and measurement. In R. Nelson (Ed.), *The rate and direction of inventive activity*. New Jersey: Princeton University Press.

Larkey, L. (1999). A patent search and classification system. In *Proceedings of the fourth ACM conference* (pp. 179–187).

Lee, S., Yoon, B., Lee, C., & Park, J. (2009). Business planning based on technological capabilities: Patent analysis for technology-driven roadmapping. *Technological Forecasting and Social Change, 76*(6), 769–786.

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

Lerner, J. (1994). The importance of patent scope: An empirical analysis. *RAND Journal of Economics, 25*(2), 319–333.

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

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

Liu, S. (2003). A route to a strategic intelligence of industrial competitiveness. In *Proceedings of the first Asia-Pacific conference on patent maps* (pp. 2–13).

Meyer, M. (2000). Patent citations in a novel field of technology – What can they tell about interactions between emerging communities of science and technology? *Scientometrics, 48*(2), 151–178.

Morris, S., DeYong, C., Wu, Z., Salman, S., & Yemenu, D. (2002). DIVA: A visualization system for exploring documents databases for technology forecasting. *Computers and Industrial Engineering, 34*(4), 841–862.

Narin, F., Noma, E., & Perry, R. (1987). Patents as indicators of corporate technological strength. *Research Policy, 16*(2–4), 143–155.

No, H., & Park, Y. (2010). Trajectory patterns of technology fusion: Trend analysis and taxonomical grouping in nanobiotechnology. *Technological Forecasting and Social Change, 77*(1), 63–75.

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

Smith, B. (1998). Optics for photolithography. In J. Sheats & B. Smith (Eds.), *Microlithography science and technology* (pp. 263–264). New York: Marcel Dekker.

Soo, V.-W., Lin, S.-Y., Yang, S.-Y., Lin, S.-N., & Cheng, S.-L. (2006). A cooperative multi-agent platform for invention based on patent document analysis and ontology. *Expert Systems with Applications, 31*(4), 766–775.

Stulen, R. H., & Sweeney, D. W. (1999). Extreme ultraviolet lithography. *IEEE Journal of Quantum Electronics, 35*(5), 694–699.

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

Tseng, Y., Lin, C., & Lin, Y. (2007). Text mining for patent map analysis. *Information Processing and Management, 43*(5), 1216–1247.

Tseng, Y., Wang, Y., Lin, Y., Lin, C., & Juang, D. (2007). Patent surrogate extraction and evaluation in the context of patent mapping. *Journal of Information Science, 33*(6), 718–736.

Verbeek, A., Debackere, K., Luwel, M., Andries, P., Zimmermann, E., & Deleus, F. (2002). Linking science to technology: Using bibliographic references in patents to build linkage schemes. *Scientometrics, 54*(3), 399–420.

Wartburg, I., Teichert, T., & Rost, K. (2005). Inventive progress measured by multi-stage patent citation analysis. *Research Policy, 34*(10), 1591–1607.

Yang, J., & Zhang, B. (2001). Customer data mining and visualization by generative topographic mapping methods. In *Proceedings international workshop on visual data mining, 4 September, Freiburg, Germany*.

Yoon, B., & 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., & Park, Y. (2005). A systematic approach for identifying technology opportunities: Keyword-based morphology analysis. *Technological Forecasting and Social Change, 72*(2), 145–160.

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