

Article

Identification of Vacant and Emerging Technologies in Smart Mobility Through the GTM-Based Patent Map Development

Jiwon Yu ¹^[b], Jong-Gyu Hwang ²,*, Jumi Hwang ¹, Sung Chan Jun ¹^[b], Sumin Kang ³, Chulung Lee ⁴^[b] and Hyundong Kim ⁵^[b]

- ¹ Department of Industrial Management Engineering, Korea University, 145 Anam-ro, Seongbuk, Seoul 02841, Korea; vermouth28@korea.ac.kr (J.Y.); hwangj0m@korea.ac.kr (J.H.); kaikai95@korea.ac.kr (S.C.J.)
- ² Center of New Transportation Innovation Research, Korea Railroad Research Institute, 176 Cheoldobangmulgwan-ro, Uiwang 16105, Gyeonggi, Korea
- ³ Department of Aerospace Engineering, University of Michigan, 500 S. State Street, Ann Arbor, MI 48109, USA; sumink@umich.edu
- ⁴ School of Industrial Management Engineering, Korea University, 145 Anam-ro, Seongbuk, Seoul 02841, Korea; leecu@korea.ac.kr
- ⁵ Division of Interdisciplinary Studies for Creativity, Daejin University, 1007 Hoguk-ro, Pocheon-si 11159, Gyeonggi-do, Korea; seanhyundongkim@gmail.com
- * Correspondence: jghwang@krri.re.kr; Tel.: +82-31-460-5438

Received: 24 September 2020; Accepted: 5 November 2020; Published: 10 November 2020



Abstract: With the development of the online platforms and the Internet of Things (IoT), various transportation services have been provided, and the lifestyle of the general public has changed significantly. However, the speed of development of technologies and services for the mobility handicapped has been relatively slow. Accordingly, in this paper, the smart mobility patent data for the mobility handicapped is subdivided through clustering to derive the mobility handicapped-related vacant technologies, and the prospect of the vacant technology is verified. For each cluster, a technology level map is generated in consideration of the technology growth level and the scope of authority of the vacant technology derived through the generative topographic map (GTM) patent map, and the level of the vacant technology is checked in terms of quantity and quality. Both indicators perform time series analyses on superior technology to predict technology trends and determine the technology's promisingness. Unlike the precedent studies that focused only on quantitative analysis methods, this paper identified the usefulness of the technology through clustering and various verification processes and materialized it as a vacant technology that is applicable to actual R&D. Accordingly, through this empirical paper, it is possible to understand the current level of vacant technology in smart mobility for the mobility handicapped and establish an R&D strategy to prevent monopoly in technology in the future market and maintain competitiveness. It can also be utilized for new technology development in consideration of convergence with currently developed technology.

Keywords: smart mobility; patent analysis; vacant technologies; mobility impairment; promising technologies

1. Introduction

Today, online platforms and the Internet of Things (IoT) are applied to mobility such as automobiles, trains, and airplanes to form closely connected transportation networks around the world. In this



paper, technologically sophisticated, fast, and eco-friendly mobility is defined as smart mobility [1,2]. Smart mobility has enriched the living environment of most people. In particular, it is contributing to the improvement of transportation convenience for mobility impairment such as the elderly, pregnant women, infants and children. Smart mobility can be more effective for mobility impairment who have a lot of difficulties in performing natural activities compared to the public. Accordingly, it is necessary to focus on the technology development of smart mobility for mobility impairment. However, it is not easy to carry out research and development (R&D) since technology development is progressing very rapidly and future technologies are unpredictable.

In order to plan R&D for smart mobility technology, technology trend analysis or technology prediction should be preceded to understand the smart mobility industry and decide the direction of the technology development [3,4]. For this, patent data have been used to obtain technical insight since patents are a final result of R&D [5,6]. Patents contain important technical information such as title of invention, summary, claims, and bibliographic information. Textual information (e.g., title, summary, claims of invention) can be used to extract new ideas that can apply to new technical R&D plans [7]. Bibliographic information such as patent number, date of patent, and applicant can be used to analyze technology trend. For this reason, patents have been utilized as an important source in analyzing rapidly evolving technology trends.

In the future, the field of smart mobility technology for mobility impaired people will be innovated more and more rapidly. As technology has advanced, competition has intensified in the market. Therefore, we attempt to empirically analyze using patent information to extract meaningful insights for predicting new promising technology that can be preoccupied in the future market. Technologies with high promisingness can be regarded as direction indicators to plan and evaluate technology development.

2. Literature Review

As deigned by the World Intellectual Property Organization (WIPO), a patent is an exclusive right granted for an invention corresponding to products or a processes that provides a new method or technical solution to solve the problems of the prior art. In order to patent, it is required to comply with standards such as originality and creativity. Many researchers spend a lot of time and effort on technology R&D, and as a result, patents are authorized as the output of R&D. Patent documents have been regarded as ample source including abundant technical information, therefore, patent-based analysis has been applied to identify the trend of technological change over the times and extract ideas for developing new technology in near future [8–10].

Patent analysis is effective when accurate results are derived [11]. The patent analysis methods such as deducting and searching for a suitable search formula, clustering, abstraction, visualization, and result-interpretation are frequently used [12]. In order to implement decision making in each process, domain expertise is required. However, if an expert does have incomplete knowledge and skills, the analysis will take a long time and the result will also be implausible. As technology advances, the number of patents is increasing faster than in the past. So the process of analyzing documents and deriving meaningful results using data mining such as text mining rather than relying only on experts is rising. Text mining is a method of extracting the essential keywords in the patent documents and generating keyword vectors that can be applied to various types of analysis such as trend analysis, technology prediction, strategic technology planning, and infringement analysis [13,14]. Consequently, a keyword vector can be generated since each keyword corresponds to a dimension. Therefore, it is possible to explain patent documents with a keyword vector.

In the previous studies, many researchers employed patent documents in forecasting vacant technologies. The methods used include patent maps [15,16], deriving vacant technologies through patent maps [17–21], patent networks [22,23], and keyword-based networks [24]. Text mining extracts subject-action-object (SAO) structure sentences that explicitly include the technical objectives of patents from patent data, and constructs a patent map to visualize the overall relationship between patents

in technology and industry. There is also a way to define potential technologies in the future by identifying vacant technologies [25]. Vacant technology can be grasped through patent map by visualizing patent documents in the multi-dimensional space into low-dimensional space. The patent map can visualize blank areas and occupied areas by analyzing patent documents based on the keyword vector. Blank areas in the map are regarded as vacant technology fields. Jun et al. (2012) made predictions of vacant technology using matrix maps and K-medoids clustering based on patent documents [26]. The top five keywords are firstly extracted to define the cluster and then identify the vacant technology area in the constructed matrix map.

Patent maps are easy to use for deriving vacant technologies. Principal component analysis (PCA), self-organizing maps (SOM), and GTM are the representative methods in constructing patent maps to identify vacant technology fields. PCA was first proposed by Pearson (1901) [27] and formulated by Hotelling (1933) [28]. It is a dimensional reduction method that converts highly correlated samples of high-dimensional space into low-dimensional. It is prediction model which works by minimizing the loss of information or characteristics of the data when reducing the dimension. Lee et al. (2009) used text mining to extract keyword vectors from patent documents and applied PCA to create a patent map [17]. Trappey et al. (2012) determined a patent with high potential value through the kaiser-meyer-olkin (KMO) approach, combining PCA and back propagation networks (BPN) [29].

SOM is a method proposed by Kohonen (1990) to perform dimensionality reduction and clustering at the same time. It is difficult to analyze data due to complicated relationships between variables through conventional PCA and other methods through nonlinear descriptive methods. It was defined as more useful for analysis [18,19]. The SOM method is superior to other multivariate approaches because it can process noisy, irregular, or missing data. PCA analyzes using only a few key components and analyzes only a portion of the total space, but SOM analyzes simply and intuitively, transforms the exact similarity values, and uses all available spaces to be more useful for analysis. Wu et al. (2016) classified patents for solar cells through the SOM, kernel PCA (KPCA), and support vector machines (SVM) method [30]. Huang et al. (2008) analyzed various chinese patent structures and clustered them by subject through the SOM method [31]. Segev and Kantola (2012) clustered the patents extracted based on the United States Patent and Trademarks Office (USPTO)'s 17 topics by applying the term frequency-inverse document frequency (TF-IDF) and SOM methods and analyzed that SOM was more effective through comparison with K-Means and density-based spatial clustering of applications with noise (DBSCAN) methods [32].

However, the SOM method does not include a cost function and lacks the rationale for selecting adjacent parameters [33]. Bishop et al. (1998) proposed a GTM method, a nonlinear latent variable model that can determine the parameters of a model using the expectation-maximization algorithm (EM algorithm) [34]. Therefore, GTM-based patent map can quantitatively derive vacant technology. GTM analyzes keyword vectors according to patent document and displays the patent document in latent grid points. In the grid, data classified through latent variables create one node, and by identifying the generated node, empty space in the grid can be detected. Each data point is mapped based on their characteristics. The nodes created at this time have the characteristics of the data that make up the node. Corresponding nodes can be extracted through the inverse function, and the part marked as a blank node can be identified. Unlike PCA that cannot detect blank areas in the map, GTM-based maps can detect blank areas automatically since latent grid points represent occupied areas and unoccupied areas by object. Therefore, it is possible to detect vacant technology fields through the GTM-based patent map. A GTM-based patent map can also interpret the vacant technology fields since GTM can be reversed from low-dimensional space into multi-dimensional space. That is, patent documents in the latent grid points in the map can be transformed into keyword vectors. GTM-based patent maps can interpret the vacant technology field in the patent vacuums, whereas SOM-based patent maps can only identify what is the patent vacuums [20,21]. Yoon and Magee (2018) identified vacant areas by applying GTM to patents visualized in a two-dimensional space, focusing on the detailed direction

4 of 22

of technology development [35]. Wu et al. (2018) compared the similarity of patent and standard documents by applying latent dirichlet allocation (LDA) and GTM [36].

GTM is an easy way to grasp vacant technology through the derived map. However, it is impossible to identify the promisingness of the vacant technologies estimated in the blank area in the patent map. Therefere, it is necessary to verify whether vacant technologies are promising. Time series analysis and technology level maps can be applied to identify promisingness of vacant technologies. Time series analysis can predict the direction whether the technology will be in the future. The technology level map is a method to identify the level of technology by considering the quantitative and qualitative levels [20,37]. Two methods are applied to analyze vacant technologies and their promisingness and through process improvement, the risk of misunderstanding is minimized. In this paper, vacant technologies are extracted using patent maps, and furthermore, the feasibility of the analyzed technologies is improved by confirming their promising properties. Applying the corresponding methodology, this paper attempts to predict a promising vacant technology in the field of smart mobility technology for mobility impairments. It proposes a technology field that is easy to enter through the analyzed promising technology, and can help in establishing an R&D strategy that can participate in new markets and strengthen technology competitiveness.

3. Methodology

3.1. Research Framework

Figure 1 shows the research framework to identify vacant technologies. Firstly, patents are collected from patent databases that is equipped in several countries such as Korea, US and China. The collected patent data cannot be used directly for analysis because patent documents include unstructured data. Therefore, preprocessing of patent data is required based on text mining. Consequently, the patent data can be converted into structured data that can be applied to patent analysis. Secondly, the preprocessed patent data is clustered by employing the K-medoids that can group the patent documents according to the same subject. Thirdly, specific keywords for each cluster are extracted by applying the TF-IDF algorithm that can evaluate the importance of words in the patent documents. Keyword vectors that can explain the feature of each patent document can be constructed. Fourthly, a GTM-based patent map is generated based on the keyword vectors. Based on the patented vacuum in the map, vacant technologies can be identified by inverse mapping from vacant areas to new vector spaces. Finally, vacant technologies derived from the GTM-based patent map are verified by applying the technology level map and ARIMA for each cluster. Therefore, we can discover the vacant technology and identify promising technologies in the field of smart mobility.



Figure 1. Research framework.

3.2. Data Collection and Preprocessing

As the competitive market becomes global, many countries and companies protect their rights by applying patents in many countries. In this paper, patent data are extracted through patent databases

of USPTO, European Patent Office (EPO), Korean Intellectual Property Office (KIPO), Japanese Patent Office (JPO), and State Intellectual Property Office of the People's Republic of China (SIPO) within WIPS ON databases. To remove duplicate patents, patent data are extracted with the patent family removed. A valid patent selection process is performed based on the extracted patent, and then the selected valid patent is analyzed. The patent data comprise largely technical content and bibliographic information. Specifically, the former has summaries and claims, the latter has IPC, Cooperative Patent Classification (CPC), and application number. In this paper, only the summary is extracted and used for analysis. As mentioned above, the patent data are unstructured data. The keywords must be extracted to analysis patent data. There are many keywords in patent documents that are not related to technology, such as dates, articles, investigations, and conjunctions. In order to derive keywords related to patent technology, preprocessing is required by removing the above stopwords and converting the abbreviation expression [33]. In this paper, the following preprocessing is performed through R programming. First, tokenizing is performed to decompose the original data in sentence form into individual words. Tokenizing refers to a series of processes that decompose words and undergo case transformation and remove spaces, numbers, and other symbols. Second, stopwords were removed through filtering. This process was conducted with experts to increase data reliability. Finally, prefixes and suffixes are removed to form a document-term matrix.

3.3. Clustering

To find the area of vacant technology using information that has been cleaned through data preprocessing, data with the same subject must be grouped. To this end, this paper applies a clustering process to the preprocessed data. Clustering is a method of analyzing unlabeled data and classifying it among data with similar properties [38]. In general, the K-means method and the K-medoids method are commonly used, and in this paper, the K-medoids method, which is less sensitive to outliers, was applied [39,40].

K-medoids are not a method of finding the center point of an object as a representative value of a cluster, but selecting a representative object that can represent a cluster form among objects. The remaining objects that have not been selected as representatives are assigned to the cluster and the representative object located closest to their location. This partitioning algorithm follows the principle of minimizing the sum of the discrepancy values and the difference between each object *i* and the representative object of the group to which *i* belongs is determined based on the absolute criterion value.

The K-medoids clustering method uses the Partitioning Around Medoids (PAM) algorithm to find the center of gravity of the cluster [41,42]. Unlike the K-means clustering method, which uses the average of K data, other distance functions such as Euclidean distance are also used to prevent data distribution distortion due to outliers. The K-medoids algorithm has a high computational cost; therefore, the optimal number of clusters is determined by applying the silhouette technique as a clustering evaluation scale [43]. Optimal number of clusters are computed as shown in Equation (1),

$$\sum_{i=1}^{k} \sum_{j=1}^{m} \|x_j^{(i)} - C_i\|^2 \tag{1}$$

where

k = number of clusters

m = number of objects

 $x_j = \text{object } j$

 c_i = centroid of cluster *i*.

Methods for measuring whether clustering is performed well can be primarily divided into internal evaluation and external evaluation. Internal evaluation is a method of evaluating data as a

result of clustering itself. It analyzes by using the cluster's similarity and gives a high score to the result with a low similarity between the clusters. Among them, the method is to evaluate clustering using random data to evaluate the non-uniformity of the data with the data in the cluster in which it belongs, and with the data in the cluster in which it does not belong as a silhouette technique. The silhouette technique indicates whether the data were clustered well, with a value between -1 and 1, indicating that the closer the silhouette value is to 1, the better the clustering is, and the closer the value is to -1, the more incorrectly classified are the data [44]. For clustering, a text-mining technique was used to extract only the part of 'summary' from the target patent to create a document–word matrix.

3.4. Keyword Extraction

The data separated through the clustering process form each group. To analyze the subject of the data these groups have and extract the essential keywords, the vacant technology for each subject has to be analyzed. The process of extracting essential keywords is used to identify the properties of each data point using the data. Keyword extraction is a technique used to extract the critical words that have to identify the attributes that the document or other data have in the field of text mining. In general, a TF-IDF weight model can be used to evaluate the importance of words existing in a document [45]. The TF-IDF model is a principle for expressing documents for information retrieval based on the vector space model. It is possible to express the relative importance of each word in individual documents, so it can be used to derive the top keywords for each cluster, and based on this, keywords included in each patent literature, such as Table 1, are generated as 1 or 0 to generate keyword vectors.

	Keyword 1	Keyword 2	 Keyword n – 1	Keyword n
Patent 1	1	0	 1	1
Patent 2	1	1	 1	1
÷	0	0	 0	0
Patent m	0	0	 0	1

Table 1. Example of keyword vector.

3.5. Identification of Vacant Technology

Using the generated keyword vector, one can analyze patent data including each keyword and identify vacant technology by identifying the characteristics of the data. In this paper, a patent map analysis method was applied to understand vacant technology. A patent map is a representative tool used as a useful measure for technical power. The patent map can easily derive useful information in the patent documents because it visualizes not only dense areas but also blank areas. Therefore, we can detect patent vacuums and interpret the complex technological properties in the patent. In the previous studies, what was drawn as a blank area on the map was supposed by patent vacuums, and each blank area was defined as a new technology and opportunity area to derive vacant technology. As a method of analyzing vacant technology, PCA, SOM, and GTM were widely used. In particular, GTM can detect and interpret patent vacuums in an automatic and objective manner [46].

GTM analyzes each patent's keyword vectors based on GTM parameters, such as the number of basis functions and late points, and maps the patent into a two-dimensional space [47–49]. Consequently, if there is a patent in each latent, it is expressed as "o," and the relatively non-patented latent is expressed as a vacant, and through Figure 2, reverse keyword vectors of the blank area are generated. As shown in Figure 2, the keyword vectors corresponding to blank nodes are calculated inversely. Applying Equation (2) in this process can yield the value of reverse keyword vectors [50,51]. Through the

extracted keyword vector, you can understand the skills of the node. This allows you to analyze what technology a node marked as vacant technology has.

$$y(s;W) = W\varphi(s) \tag{2}$$

where

W = initial weighting matrix

 $\varphi(s)$ = the activation of basis functions for latent variable *s*

y(s; W) = transforms the latent point x on the map.



	keyword 1	keyword 2	keyword 3	keyword 4	keyword 5	keyword 6	 keyword N
node 1	0.04469	0.12797	0.49404	0.23475	0.39398	0.87220	0.66154
node 2	0.22031	0.07647	0.38522	0.09177	0.48062	0.45332	0.54911
node 3	0.08860	0.59250	0.06740	0.48668	0.95157	0.69737	0.00860
node 4	0.16666	0.03984	0.16183	0.31873	0.17489	0.55192	0.48073
node 5	0.19830	0.35401	0.07773	0.18368	0.15931	0.37780	0.28516
node 6	0.66683	0.51172	0 56008	0.99039	0.21692	0.14814	0.57543
node 7	0.86539	0.01582	0.88574	0.49143	0.00130	0.01145	0.59549
node K X K	0.61294	0.45970	0.11283	0.97069	0.28263	0.26450	0.97330

Figure 2. Example of patent-vacant technology.

3.6. Verification of Vacant Technology

The process of deriving blank nodes from the blank area using the GTM algorithm is a quantitative method. However, it is necessary to verify whether the keywords are promising because the derived blank keywords do not have viable opportunities for technology development [35]. Accordingly, this paper validates the vacant technology level derived from the blank node and predicts the promising vacant technology through time series analysis.

3.6.1. Identification of Vacant Technology Level

The blank node derived through GTM is only a macro-level type of information which means that the patent vacuums are technological opportunities, so it is not enough to be used as a decision-making tool for planning and carrying out actual technology development. In addition, the vacant technology keyword derived from the blank node can more specifically confirm the result with textual information, but it is difficult to confirm the presence or absence of actual value. Accordingly, verification of the level of blanking techniques derived is necessary.

Yoon et al. (2019) confirmed the level of technology considering the patents' quantitative and qualitative levels. In this paper, the quantity standard was evaluated by the number of patents, and the quality standard was evaluated by the citation frequency [20]. However, the characteristic of this paper is that the quality of patents is defined based on the number of patent vacuums claims because the frequency of citations is low or high. The number of claims refers to the scope of the legal authority of the patent [17], which can be interpreted as a patent with higher quality because the higher the number of claims, the higher the technological competitiveness and monopoly authority.

The quantity of patent can be calculated as Equation (3) using the concept of revealed technological advantage (RTA). The quantity is the ratio of the frequency of patents containing keywords derived from vacant technology in patent by cluster, which means the level of growth of the vacant technology in the cluster. The quantitative level of the vacant technology is determined by the gap between the vacant technology and the other vacant technology. It is calculated by substracting the RTA average of

the other vacant technologies from the RTA value of the target vacant technologies. It is calculated by substracting the RTA average of vacant technologies. Subsequently, the quality of the patent uses the scope of technology (SoT) index Equation (4), which measures the average number of patent claims containing keywords derived from the cluster-specific patent gap. Consequently, the qualitative level of each vacant technology is calculated by substracting the average of the SoT values of other vacant technologies from the SoT value of the target vacant technology. As a result, the technical level of the derived vacant technologies was defined as shown in Figure 3.

$$Technology growth = \frac{N_i}{N}$$
(3)

where

N = the number of total patents in each Clusters N_i = the number of patents including keyword *i*.

$$\Gamma echnology \ scope \ = \ \frac{C_i}{C} \tag{4}$$

where

C = the number of total claims in each Clusters

 C_i = the number of claims of patents including keyword *i*.





In this paper, quantitative/qualitative superiority technology (①) in Figure 3, which is considered promising due to its high quantitative and qualitative level, is used as an analysis target, and the corresponding technology is extracted, and the promising technology is predicted through time series analysis.

3.6.2. Prediction of Vacant Technological Promisingness

Promisingness of technology has properties such as radical novelty, relatively fast growth, coherence, noticeable impact, and uncertainty [37]. In particular, the growth rate of technology can be confirmed through the trend of increasing the frequency of patent filings by year, conducting autoregressive integrated moving average (ARIMA) analysis by time series data of patent filing frequency by year, and promisingness of technology through trends derived from the analysis results. The ARIMA model is combined (Equation (7)) of the auto-regression (AR) model (Equation (5)) and the moving average (MA) model (Equation (6)). AR model (Equation (5)) is a model that constructs autocorrelation as a time series model using past values, and the MA model (Equation (6)) is a model that predicts a specific point in the future based on past time series data. Using the ARIMA model (Equation (7)) is suitable for predicting the future using past data [52].

The future trend of the keyword can be predicted through ARIMA analysis, and based on this, keywords that are likely to become mainstream in the future market can be identified [53].

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p}, \varepsilon_t \sim i.i.d, N(0, \sigma^2)$$
(5)

where

 α_i = the parameters of the autoregressive part

 Y_t = lagged values:

$$Y_t = e_t - \beta_1 e_{t-1} - \beta_2 e_{t-2} - \dots - \beta_q e_{t-q}$$
(6)

 β_q = moving average coefficient e_t = error term at time t:

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t, \tag{7}$$

where

 ε_t = white noise.

4. Deriving Promising Technologies through Patent Analysis: Smart Mobility

4.1. Data Collection and Preprocessing in Smart Mobility

The technical research is conducted as preliminary research in the development process, and its level can be confirmed through the thesis; research related to ICT smart mobility for mobility handicap was selected by searching for keywords needed for patent search formulas. Based on the searched 20 papers, 40 keywords such as *Artificial Landmark Recognition*, *Bluetooth*, *Collaborative*, *Fuzzy*, *Geographic Information*, *and Infrared Sensor* were extracted. Based on these keywords, the search period in USPTO, EPO, JPO, SIPO, and KIPO database was set to 1990.01.01 to 2020.05.23 to conduct a patent search. The patent data collection gradually expanded the search formula by adding new keywords from patents extracted based on initial keywords. The final expanded keywords are shown in Appendix A. As a result of searching for patents based on the final keywords, 5210 patents were searched. The patents were collected by ten experts in the field of smart mobility technology to select valid patents related to this paper; as a result of collecting and analyzing the experts' judgments, 2619 valid patents for smart mobility were selected.

For the preprocessing of training data, a document–keyword matrix was created by extracting only the *summary* part of the target patent using the text-mining technique. Since it is challenging to interpret multidimensional variables structurally, it is analyzed by converting them to low dimensions through principal component analysis. In this paper, the number of principal components for each country's patent document is selected so that the principal components' explanatory power exceeds 90 percent.

In addition, the misspelled words or abbreviations included in the extracted data were corrected or removed [47]. The original data in sentence format were decomposed into each word and made lowercase, and spaces, numbers, and other symbols were removed through the tokenizing process. Subsequently, through filtering, the words that were deemed unnecessary, the prefixes and suffixes were removed, leaving only the words that formed the document, therefore forming a patent-keyword Matrix, such as Table 2. Currently, the number of keywords was too large to use all the extracted keywords; therefore, only the top 500 were extracted to form a 500 by 2619 matrix.

		Patent 1	Patent 2	Patent 3	•••	Patent 2619
1	tag	1	0	1		1
2	bluetooth	1	1	1		1
3	vehicle	0	0	1		0
4	location	0	0	1	•••	1
÷	:	0	1	0		0
500	data	1	1	1	•••	0

Table 2. Patent-Keyword Matrix in ICT Smart Mobility.

4.2. Clustering from Smart Mobility Patent

The K-medoids clustering method was used to determine the optimal number of clusters. In determining the optimal number of clusters, the distribution of data is likely to be distorted due to anomaly if clustering is carried out through analysis of principal components, so clustering is carried out using the K-medoids method, which is robust to ideal values.

In clustering, the silhouette analysis was applied to indicate how well the data were clustered, and when determining the optimal number of clusters, the number of clusters with a silhouette value closest to 1 was chosen [45]. As shown in Figure 4, this analysis shows that if there are three clusters, the silhouette value is the closest cluster to 1, so the optimal number of clusters is chosen as three.



Figure 4. Determining the optimal number of clusters.

As a result of clustering, the number of patents corresponding to each cluster consists of 919 for cluster 1, 1149 for cluster 2, and 551 for cluster 3. GTM was applied to each cluster to understand the vacant technology of each cluster.

4.3. Keyword Extraction from Smart Mobility Patent

GTM is a method used to identify vacant technologies for specific technologies using patent maps. A keyword–patent matrix was created by extracting keywords from the data to create a patent map. In this paper, a keyword–patent matrix was created by extracting the keywords of the clustered groups to grasp the vacant technology of each clustered group. The keyword vector is composed through the keyword frequency and TF-IDF weight of the document [54]. To create a keyword–patent matrix for each cluster and use that matrix to create a patent map, it is recommended to match the number of keywords in the matrix equally. Among the tf-idf values that can be extracted equally, 0.5, the highest value, is applied as a standard to enhance explanatory power and enable key technical keywords to be extracted. As a result, 40 keywords were selected for each group. Each keyword represents a property of the vector; the keyword vector has 40 dimensions. The purpose of this paper is to judge the vacant technology and to judge the promising technology for that vacant technology, so only whether the patents currently have keywords is necessary. Accordingly, we used the binary keyword vector, not the frequency keyword vector [37]. The keyword–patent matrix indicates whether each patent contains a keyword. Each patent contains 1 if the keyword is included, otherwise 0.

4.4. Vacant Technology in Smart Mobility

4.4.1. GTM Performance Results

Based on the completed keyword–patent matrix, the GTM algorithm is applied to derive vacant technology. The number of nodes (K) is implemented as a variable for GTM map creation. According to Son et al. (2012), who conducted the patent analysis based on the GTM-based patent map, an appropriate value of K should be assigned so that the map cannot be too sparse or dense [46]. Therefore, experiments were conducted within the K range of 10 to 17. The value of 14 was judged to be the most appropriate, consequently, a 14 × 14 map was developed. The GTM-based map was derived for Cluster 1–3. It should be derived from the vacant technology through the corresponding patents and keywords, identifying the blank node in the GTM map to verify the patent vacuums. Figure 5 shows a blank node based on the GTM map of Clusters 1–3.



Figure 5. Blank node: (a) Cluster 1, (b) Cluster 2, (c) Cluster 3.

The GTM map derived 10 blank nodes for Cluster 1, 7 for Cluster 2, and 28 for Cluster 3. The blank nodes displayed in the GTM map can derive patents and keywords owned by the node through an inverse function. Table 3 shows patents and keywords for blank nodes in each cluster. Table 4 shows the keywords that should be finally checked by removing duplicate words.

Cluster	Blank Node	Keyword
1	1	sensor, vehicle, block, tag, digital
	6	intelligent
	9	intelligent, sensor, vehicle, response, computer, radio
	10	location, intelligent, positioning, response, memory, computer
2	7	bluetooth, location, network
	1	tag, block, data, wireless
	2	tag, beacon, block
	3	beacon, block, data, label
	4	block
3	6	electronic, rfid
5	8	tag, beacon, label, antenna, wireless
	9	bluetooth, control
	11	tag, block, electronic
	25	bluetooth, electronic, reader
	28	tag, beacon, block, data

Table 3. Keywords belonging to the empty nodes of Cluster 1, 2, and 3¹.

¹ No patent exists in the blank node other than the blank node above.

Cluster	Keyword
1	block, computer, digital, intelligent, location, memory, positioning,
1	radio, response, sensor, tag, vehicle
2	bluetooth, location, network
3	antenna, beacon, block, bluetooth, control, data, electronic, label, reader,
5	rfid, tag, wireless

Table 4. Vacant technology keywords for each cluster
--

An analysis of the node-marked vacant technology shows that it is divided into nodes with and without keywords. Keywords are extracted if a patent is associated with the vacant technology corresponding to the node, but not if there is no relevant patent. The absence of related patents means that the technology corresponding to the node has not been studied. In other words, the related patent has not been issued, and the keyword cannot be extracted as there is no corresponding patent [54].

4.4.2. SOM Performance Results

For each cluster, the results of the SOM were derived for comparison with the GTM. Figure 6 shows a blank node based on the SOM map of Clusters 1–3. The nodes displayed in gray on the SOM map are blank nodes. SOM map derived 17 blank nodes for Cluster 1, 3 for Cluster 2, and 27 for Cluster 3.



Figure 6. Blank node: (a) Cluster 1, (b) Cluster 2, (c) Cluster 3.

In the patent map using SOM, there is a limit to analyzing nodes currently marked as blank nodes. As shown in Figure 7, the node in the SOM map consists of the number of the patent corresponding to each node. The patent corresponding to that node does not exist, it is displayed as a blank node. In this case, the blank node cannot be interpreted because there is no patent data that makes up that node. In this case, the empty node cannot be interpreted using the inverse function, because there are no patent data constituting the node.



Figure 7. Example of self-organizing maps (SOM) map.

To interpret the blank node, identify the patents held by the nodes surrounding the blank node. The common keywords of the corresponding patents are linked to the technology of the blank node. Therefore, it extracts keywords of surrounding patents and infers blank technology through it. In this paper, keywords in the top 20% were defined as common keywords by analyzing keywords of surrounding patents and identifying the frequency of their appearance. This is because it was judged to be difficult to analyze due to the low frequency and small number of related patents in the case of the top 20% or less.

In addition, if there is a series of vacuum technologies when performing the analysis, the analysis was grouped together into one vacuum technology because the neighboring nodes were shared and the nearby technologies were analyzed. An analysis of the node-marked vacant technology shows that it is divided into nodes with and without keywords. Table 5 shows patents and keywords for blank nodes in each cluster. Table 6 shows the keywords that should be finally checked by removing duplicate words.

Cluster	Blank Node	Keyword
	16	system, intelligent
	23	response
	46	intelligent, vehicle, sensor
	67,68	positioning
	83	positioning, robot
	93	transmitter
1	95	transmitter
1	119	data
	121	transmitter
	123,137,138	transmitter
	150	location, detection, vehicle, transmitter
	165	location, detection, transmitter, identification
	178	tag, location, transmitter, detection
	181	location, robot, detection
	49	Bluetooth, location
2	67	Application
	90	Bluetooth, location, network
	17	access, connection
	27	antenna, beacon, data, tag
	30	connection, reader
	51	unit
	55	antenna, tag
	96,97	antenna
3	119,120,121	tag
	130	beacon, data
	137	invention, information
	141,142	beacon, tag, data, multiple
	147	tag
	170	light, data, beacon
	179	storage, signal, bluetooth

Table 5. Keywords belonging to the empty nodes of Cluster 1, 2, and 3¹.

¹ No patent exists in the blank node other than the blank node above.

Cluster	Keyword
1	data, detection, identification, intelligent, location, positioning, response, robot, sensor, system, tag, transmitter, vehicle
2	application, bluetooth, location, network
3	access, antenna, beacon, bluetooth, connection, data, information, invention, light, multiple, reader, signal, storage, tag, unit

4.4.3. Comparative Analysis with GTM and SOM

Tables 4 and 6 analyze vacant technology keywords derived from SOM and GTM as follows. Analyzed with SOM, the vacant technology keywords derived are 13, 4, and 15, respectively. Analyzed with GTM, the keywords for vacant technology are derived from 12, 3, and 12, respectively. This is shown in Table 7. SOM was derived more by analyzing only the total number of keywords, but this is a keyword analyzed through patents constituting the neighboring nodes, and it is difficult to analyze as a keyword of a vacant technology that describes the cluster or node. The overall number of keywords in GTM is small, but it contains more essential content. This is because to interpret a blank node in SOM, the analyzed vacant technology is inaccurate because the patent constituting the nodes around the blank node must be identified and inferred through it. Furthremore, the blank node technologies analyzed are not extracted using data from the node itself, but they are analogous to other patents around it, so one cannot analyze the promising of the technology. Moreover, the blank node technologies analyzed are not extracted using data from the node itself, but they are analogous to other patents around it, so one cannot analyze the promising of the technology.

Table 7. Number of vacant technology keywords.

Cluster	SOM	GTM
1	13	12
2	4	3
3	15	12

In GTM, blank nodes can be interpreted using the inverse function. Each blank node has the advantage of being more useful in terms of vacant technology analysis than SOM in that it can analyze its technology more accurately and confirm its promisingness; GTM is applied to this study.

4.5. Verification of Vacant Technology in Smart Mobility

4.5.1. Identification of Vacant Technology Level

The current technology level must be checked for patents corresponding to keywords derived through GTM. Figures 8–10 show the graphs of the vacant technology level for each cluster based on the number of patent citations and the number of patent claims.



Figure 8. Status of Cluster 1's vacant technology level.



Figure 9. Status of Cluster 2's vacant technology level.



Figure 10. Status of Cluster 3's vacant technology level.

Identifying the status of the vacant technology level in Cluster 1, six keywords (technology) corresponding to the high technology growth–high technology scope that is considered promising were derived. The keywords (technology) are *location* (10), *intelligent* (6), *sensor* (1), *vehicle* (2), *positioning* (11), and *block* (3) and a prediction needs to be made on whether they are promising technologies for the future. Conversely, *radio* (9), *digital* (5), *computer* (8), *memory* (12), and *tag*(4) are some of the keywords that fall under low technology growth–low technology scope. *Response* (7) is derived for the keyword (technology), which falls under low technology scope–high technology growth.

Identifying the status of the vacant technology level in Cluster 2, only one keyword (technology) corresponding to the high technology growth–high technology scope considered promising was derived. The keyword (technology) is *Bluetooth* (1) and it needs to be predicted whether they are promising technologies in the future. Conversely, *location* (2) and *network* (3) were derived for technologies that correspond to low technology grouting–low technology scope.

Identifying the status of the vacant technology level in Cluster 3, six keywords (technology) corresponding to the high technology growth–high technology scope considered promising were derived. The keywords (technology) are *tag* (8), *beacon* (12), *data* (10), *antenna* (4), *electronic* (2), and *block* (9) and it needs to be predicted whether they are promising technologies in the future. Conversely,

wireless (11), *control* (6), *rfid* (3), *reader* (7), and *label* (1) were derived for technologies that fall under low technology growth–low technology scope. *Bluetooth* (5) was derived for the keyword (technology), which is low technology scope–high technology growth.

4.5.2. Prediction of Vacant Technological Promisingness

Prediction of promising technologies through time series analysis to identify prospectiveness for technologies that belong to high technology growth–high technology scope among keywords was done through GTM. For the time series analysis, the ARIMA model was used, and the parameters p, q, r of the ARIMA model were estimated through R. p, q, and r represent the order of the autoregressive model, the degree of difference, and the order of the moving average model. The values of p, q, and r used the auto arima function of R, which adjusts itself so that each result is optimally displayed. Each keyword for each cluster is drawn in two types, ascending increase and decline through the ARIMA model, as shown in Table 8. This paper predicts the prospect of vacant technology, the type that rises increase is defined as a technology with prospects and schematizes the ARIMA results.



Table 8. Vacant technology prospective outcome type.

Figures 11–13 shows the ascending type by applying ARIMA to technologies belonging to the high technology growth–high technology scope. It confirms that the technologies derived from Cluster 1 ascending increase are *sensor* (1) and *vehicle* (2). Subsequently, *Bluetooth* (1) for Cluster 2, and *beacon* (12) and *data* (10) for Cluster 3 were derived.



Figure 11. Cluster 1's Type 1 technology.



Figure 12. Cluster 2's Type 1 technology.



Figure 13. Cluster 3's Type 1 technology.

4.6. Analysis Result in Smart Mobility

Looking at the analyzed promising technologies for each cluster, Cluster 1 is predicted to be about a vehicle technology that supports the mobility handicapped and a sensor technology that recognizes and analyzes surrounding objects to prepare for problems that may occur for the mobility handicapped. Cluster 2 is predicted to be a Bluetooth technology that can transmit useful information. Cluster 3 is predicted to be about the data that are useful when delivered to the mobility handicapped, and a beacon technology installed on facilities which transmits useful information to the mobility handicapped based on the analysis of their movement.

As a result of deriving blank nodes through the GTM algorithm and deriving vacant technology for each cluster, twelve vacant technologies were derived for Cluster 1, three for Cluster 2, and twelve for Cluster 3. To investigate the technology level of vacant technology, a technology growth—technology scope graph based on the number of patent applications and patent claims of the keyword was created. For the technologies that were found to be promising through the graph, six vacant technologies in Cluster 1 were found to have a high prospectiveness, and one in Cluster 2 and seven in Cluster 3 respectively. When applying the time series analysis through the ARIMA model to confirm whether the technologies are promising in the future, the terms sensor and vehicle are the technologies that are both vacant technologies and promising in Cluster 1. Respectively, for those of Cluster 2 it is Bluetooth, and for Cluster 3 they are beacon and data.

Looking at the degree of technology development of these technologies through Figures 11–13, all five promising technology keywords have fewer than ten patents applied at first, and then the number of patents increases rapidly at a certain fiducial point. Cluster 1 began to increase in the number of patents from 2010, Cluster 2 increased from 2005, and finally, Cluster 3 increased from 2008. From Cluster 2 and 3, Bluetooth, data, and Beacon technologies are analyzed to have increased in necessity and technology since the mid-2000s. Cluster 1 decreases slightly after a sudden increase and

then starts to increase again. This is because the vehicle and sensor technologies have been developed since the early 2010s, and the necessity of going through the process of decrease–increase has resulted in this graph.

As a whole, the number of patents for clusters 1–3 has increased dramatically over the last 10 to 15 years. Recently, the need for related technologies corresponding to ICT Smart Mobility is gradually increasing, and many related patents have been applied. The technology is expected to be in the current development stage based on the number of patents filed and is expected to increase steadily in the future.

As a result of the analysis, the identification process of promising, vacant technologies, and the validity of the identified technology area have improved. The promising technologies of ICT smart mobility can be analyzed through keywords, and the three promising technologies analyzed are as follows. First, it is a technology that smoothly supports the movement of the mobility handicapped by notifying the movements of nearby objects and notifying them when problems occur. Furthermore, it recognizes the movements of objects around the device and recognizes unexpected problems first, and alerts the surroundings to prepare for unexpected accidents and to support the movement of the mobility handicapped smoothly. The user's physical characteristics are recognized through recognition, the user is guided through a method suitable for the characteristics, and obstacles can be detected and avoided by additionally recognizing the surrounding environment. In particular, it searches evacuation routes in the event of a disaster and helps traffic vulnerable people who are more difficult to respond to emergencies to evacuate. Second, it is an information communication technology that delivers useful information for the mobility handicapped. It analyzes indoor and outdoor images and uses various methods to guide the route that is appropriate for pedestrians. It provides a service that selects and transmits the optimal route for each mobility handicap type. The provisioning method is also a technology that delivers useful information so that there is no inconvenience when a mobility handicapped person moves using a mobile phone or public facilities. Finally, it is a wireless positioning technology to deliver useful information so that there is no discomfort when the traffic weak moves. It stores information that is useful to the underprivileged, such as ramps, special structures, facilities, and guide information, by 3D modeling, and provides the information when the traffic underprivileged approaches. The stored device is a technology that provides information according to the situation by grasping the movement of the nearby mobility handicapped person through devices such as a cane after being installed in nearby facilities. Through the three technologies, the movement of the mobility handicapped can be easily grasped in the future, provide useful information according to the situation, and understand that the technology that can guide the route and support the movement will be developed.

5. Conclusions

In this paper, a reliable method for identifying promising vacant technology areas is provided, and patent analysis reveals the characteristics and meaningful insights into ICT smart mobility technology. In many papers, various methodologies have been applied to derive promising vacant technologies. Among them, the method using GTM was suitable for deriving vacant technologies through a patent map. In this paper, we further explored the promisingness of vacant technology and analyzed technologies with great potential for future development. As a result of the analysis, the identification process of promising, vacant technologies is improved, and the validity of the identified technology area is improved. The promising technology of ICT smart mobility analyzed through keywords is a (1) vehicle technology that smoothly supports the movement of the mobility impaired person through notification when there is a problem by grasping the movement of surrounding objects, (2) information communication technology to deliver useful information so that there is no inconvenience when the mobility impaired person moves, (3) wireless positioning technology that is installed in a facility and provides information by identifying the movement of a nearby mobility impaired person. Through the derived technologies, it can be understood that in the future, technologies that can recognize mobility handicaps, provide useful information to people

who have them, and guide paths will be developed. The population of the mobility handicapped will gradually increase, and the area is expected to continue to grow as its demand increases.

To explore opportunities, it is necessary to invent the core technology in the field or to gain a competitive edge that differentiates it from other companies or countries. In order to discover opportunities in the future market and to create ideas that will be useful, we analyzed the vacant technology area, which is considered promising, and extracted keywords that are assumed to represent promising technologies. "Sensor", "vehicle", "bluetooth", "beacon" and "data" were identified as the most superior and promising technologies. These technologies can be analyzed with technologies related to assistive devices that help mobility impaired people move and prevent problems when using mobility. Technologies related to smart devices that can recognize mobility impaired people, predict and prevent possible problems, and provide useful information will be needed, and these technologies will provide convenience to mobility impaired people. In order to preoccupy opportunities in the future market, it is necessary to characterize dangerous objects or mobility impaired people when they are detected and develop strategies related to technologies or products that can transmit the collected information. For this strategy, the methodology proposed in this paper can be applied to break down the technologies that have been applied to date and analyze which areas of technology have been applied mainly. GTM allows to create the patent map for that technology and identify areas of vacant technology that are not currently applied. The technology level map and ARIMA can then be used to determine the potential for future development. This research can provide useful information to managers in developing R&D strategies in ICT smart mobility technologies and can be sufficient data to plan new project.

The same procedure can be used to identify new business-creating potential. The proposed methodology is applied, making it easy to identify and analyze vacant technologies that are not currently developed by visualizing patent data of related technologies in a two-dimensional space. The possibility of future market development of the derived vacant technology can be grasped, and only useful information for strategy establishment can be extracted, so that the possibility of technological development can be grasped and related information can be provided more abundantly. This shortens the work time for managers and experts and improves operational efficiency.

By applying the methodology suitable for smart mobility, the vacant technology with high potential for future development beyond the limitations of existing GTM-based patent analysis was identified. Moreover, using the patent map extends from the macroscopic definition of vacant technology to provide the promising-ness of vacant technology. Prospects are centered on keywords, which can help develop a vacant technology development plan. This paper has the effect of improving the process of identifying either promising or vacant technologies and increasing the effectiveness of the identified technology areas.

The exploration of new technological opportunities will help companies gain a competitive advantage and lead to economic growth. To identify and preoccupy new technologies, it is necessary to predict promising technologies with high potential for development. The proposed method is suitable for identifying the current vacant technology and analyzing the prospects. Promising areas analyzed through the proposed method for the development of ICT smart mobility technology will continue to grow, and ICT smart mobility technology is expected to be more secure and developed. It provides more detailed information to decision-makers who need to predict promising technologies in establishing R&D strategies in ICT smart mobility technology.

Author Contributions: Conceptualization, J.-G.H., J.Y. and C.L.; methodology, J.Y., J.-G.H., J.H., S.C.J. and H.K.; validation, J.-G.H. and C.L.; formal analysis, J.Y., J.H., S.C.J. and S.K.; investigation, J.Y., J.H. and S.C.J.; data curation, J.Y.; writing—original draft preparation, J.Y., J.H., S.C.J. and S.K.; writing—review and editing, J.Y., J.-G.H., J.H., S.C.J., S.K. and C.L.; visualization, J.Y., J.H. and S.C.J.; supervision, J.-G.H. and C.L.; project administration, J.-G.H., C.L. and H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by "Development in Interactive Route Guidance and Supporting System Technology for the Mobility Handicapped in Railway Station" from the R&D Program of the Korea Railroad Research Institute, Republic of Korea.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1	. Smart	Mobility	Patent	search	keyword.
----------	---------	----------	--------	--------	----------

Division	Keyword
#	3D Modelling
А	Artificial Landmark Recognition, Auditory sensor, Automatic Speech Translation
В	Beacon, Biosensor, Bluetooth, Bluetooth Low Energy, Braille Block, Building Information Modeling
С	Child, CoBoT, Collaborative Robot, Command Recognition, Crowdsourcing
D	Dead Reckoning, Distance Sensor
F	Fuzzy
G	Geographic Information, Geomagnetism, Gesture, Grasping
Н	Heading Estimation, Human Augmentation
Ι	I/O Interface, Image Matching, Indoor Positioning, Infants Companion, Infrared sensor, IoT
L	Language Identification, Laser sensor, LBS, Localization, Location
М	Manipulator, Micro wave, Mobile App, Mobility Handicapped, Motion Recognition, Multi modal
N	Natural Landmark Recognition, NFC
0	Obstacle Avoidance, Opportunistic Signal
Р	Pedestrian Dead Reckoning, Pregnant Woman
R	Radio Mapping, Received Signal Strength Indication, Redundant, RFID
S	Senior Citizen, Simultaneous Localization Mapping, SLAM, smart mobility, Smart Stick, Sound Recognition,
5	Speech Recognition
Т	Tag, Text Speech, TTS
U	Ultrasonic sensor
V	Vision Sensor, Voice User Interface
W	Walking Stabilization

References

- Docherty, I.; Marsden, G.; Anable, J. The governance of smart mobility. *Transp. Res. Part A Policy Pract.* 2018, 115, 114–125. [CrossRef]
- 2. Zawieska, J.; Pieriegud, J. Smart city as a tool for sustainable mobility and transport decarbonisation. *Transp. Policy* **2018**, *63*, 39–50. [CrossRef]
- 3. Gerdsri, N.; Assakul, P.; Vatananan, R.S. An activity guideline for technology roadmapping implementation. *Technol. Anal. Strateg. Manag.* **2010**, *22*, 229–242. [CrossRef]
- 4. Smith, M.; Agrawal, R. A Comparison of Time Series Model Forecasting Methods on Patent Groups. *MAICS* **2015**, *1353*, 167–173.
- 5. Rivette, K.G.; Kline, D. *Rembrandts in the Attic: Unlocking the Hidden Value of Patents*; Harvard Business Press: Brighton, MA, USA, 2000.
- 6. Park, H.; Yoon, J.; Kim, K. Identification and evaluation of corporations for merger and acquisition strategies using patent information and text mining. *Scientometrics* **2013**, *97*, 883–909. [CrossRef]
- 7. Barcelon, Y.C. Role of patent analysis in corporate R&D. Pharm. Pat. Anal. 2012, 2, 5–7.
- Huang, Z.; Chen, H.; Yip, A.; Ng, G.; Guo, F.; Chen, Z.K.; Roco, M.C. Longitudinal patent analysis for nanoscale science and engineering: Country, institution and technology field. *J. Nanoparticle Res.* 2003, *5*, 333–363. [CrossRef]
- 9. Feng, X.; Fuhai, L. Patent text mining and informetric-based patent technology morphological analysis: An empirical paper. *Technol. Anal. Strateg. Manag.* **2012**, *24*, 467–479. [CrossRef]
- 10. Gao, L.; Porter, A.L.; Wang, J.; Fang, S.; Zhang, X.; Ma, T.; Wang, W.; Huang, L. Technology life cycle analysis method based on patent documents. *Technol. Forecast. Soc. Chang.* **2013**, *80*, 398–407. [CrossRef]
- 11. Madani, F.; Weber, C. The evolution of patent mining: Applying bibliometrics analysis and keyword network analysis. *World Pat. Inf.* **2016**, *46*, 32–48. [CrossRef]
- 12. Tseng, Y.H.; Lin, C.J.; Lin, Y.I. Text mining techniques for patent analysis. *Inf. Process. Manag.* 2007, 43, 1216–1247. [CrossRef]
- Abbas, A.; Zhang, L.; Khan, S.U. A literature review on the state-of-the-art in patent analysis. *World Pat. Inf.* 2014, *37*, 3–13. [CrossRef]

- 14. Noh, H.; Jo, Y.; Lee, S. Keyword selection and processing strategy for applying text mining to patent analysis. *Expert Syst. Appl.* **2015**, *42*, 4348–4360. [CrossRef]
- 15. Yoon, B.; Yoon, C.; Park, Y. On the development and application of a self–organizing feature map-based patent map. *RD Manag.* **2002**, *32*, 291–300. [CrossRef]
- 16. Yoon, J.; Kim, K. Detecting signals of new technological opportunities using semantic patent analysis and outlier detection. *Scientometrics* **2012**, *90*, 445–461. [CrossRef]
- 17. Lee, S.; Yoon, B.; Park, Y. An approach to discovering new technology opportunities: Keyword-based patent map approach. *Technovation* **2009**, *29*, 481–497. [CrossRef]
- 18. Kohonen, T. The self-organizing map. Proc. IEEE 1990, 78, 1464–1480. [CrossRef]
- 19. Juntunen, P.; Liukkonen, M.; Lehtola, M.; Hiltunen, Y. Cluster analysis by self-organizing maps: An application to the modelling of water quality in a treatment process. *Appl. Soft Comput.* **2013**, *13*, 3191–3196. [CrossRef]
- Yoon, B.; Park, I.; Yun, D.; Park, G. Exploring promising vacant technology areas in a technology-oriented company based on bibliometric analysis and visualisation. *Technol. Anal. Strateg. Manag.* 2019, *31*, 388–405. [CrossRef]
- 21. Huang, M.; Chang, H.; Chen, D. The trend of concentration in scientific research and technological innovation: A reduction of the predominant role of the US in world research & technology. *J. Informetr.* **2012**, *6*, 457–468.
- 22. Yoon, J.; Kim, K. Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks. *Scientometrics* **2011**, *88*, 213–228.
- 23. Alstott, J.; Triulzi, G.; Yan, B.; Luo, J. Mapping technology space by normalizing patent networks. *Scientometrics* **2017**, *110*, 443–479. [CrossRef]
- 24. Choi, J.; Hwang, Y.S. Patent keyword network analysis for improving technology development efficiency. *Technol. Forecast. Soc. Chang.* **2014**, *83*, 170–182. [CrossRef]
- 25. Yoon, J.; Park, H.; Kim, K. Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis. *Scientometrics* **2013**, *94*, 313–331.
- 26. Jun, S.; Park, S.; Jang, D. Technology forecasting using matrix map and patent clustering. *Ind. Manag. Data Syst.* **2012**, *112*, 786–807. [CrossRef]
- 27. Pearson, K.L. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dub. Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [CrossRef]
- Hotelling, H. Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. 1933, 24, 417–441. [CrossRef]
- 29. Trappey, A.J.C.; Trappey, C.V.; Wu, C.; Lin, C. A patent quality analysis for innovative technology and product development. *Adv. Eng. Inform.* **2012**, *26*, 26–34. [CrossRef]
- 30. Wu, J.T. A patent quality analysis and classification system using self-organizing maps with support vector machine. *Appl. Soft Comput.* **2016**, *41*, 305–316. [CrossRef]
- Huang, S.H.; Ke, H.R.; Yang, W.P. Structure clustering for Chinese patent documents. *Expert Syst. Appl.* 2008, 34, 2290–2297. [CrossRef]
- Segev, A.; Kantola, J. Identification of trends from patents using self-organizing maps. *Expert Syst. Appl.* 2012, 39, 13235–13242. [CrossRef]
- 33. Lee, C.; Kang, B.; Shin, J. Novelty-focused patent mapping for technology opportunity analysis. *Technol. Forecast. Soc. Chang.* **2015**, *90*, 355–365. [CrossRef]
- Bishop, C.M.; Svensén, M.; Williams, C.K.I. GTM: The generative topographic mapping. *Neural Comput.* 1998, 10, 215–234. [CrossRef]
- Yoon, B.; Magee, C.L. Exploring technology opportunities by visualizing patent information based on generative topographic mapping and link prediction. *Technol. Forecast. Soc. Chang.* 2018, 132, 105–117. [CrossRef]
- Wu, F.; Mi, L.; Li, X.; Huang, L.; Tong, Y. Identifying Potential Standard Essential Patents Based on Text Mining and Generative Topographic Mapping. In Proceedings of the 2018 IEEE International Symposium on Innovation and Entrepreneurship, Beijing, China, 30 March–1 April 2018; pp. 1–9.
- Kadhim, A.I.; Cheah, Y.; Ahamed, N.H. Text document preprocessing and dimension reduction techniques for text document clustering. In Proceedings of the 2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology, Kota Kinabalu, Malaysia, 3–5 December 2014; pp. 69–73.

- Madhulatha, T.S. Comparison between K-Means and K-Medoids Clustering Algorithms. In Communications in Computer and Information Science, Proceedings of the International Conference on Advances in Computing and Information Technology, Chennai, India, 15–17 July 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 472–481.
- 39. Velmurugan, T.; Santhanam, T. Computational complexity between K-means and K-medoids clustering algorithms for normal and uniform distributions of data points. *J. Comput. Sci.* **2010**, *6*, 363–368. [CrossRef]
- 40. Arora, P. Analysis of k-means and k-medoids algorithm for big data. *Proc. Comput. Sci.* **2016**, *78*, 507–512. [CrossRef]
- 41. Pyo, G.C.; Park, J.W.; Moon, S.I. A new method for dynamic reduction of power system using PAM algorithm. In Proceedings of the IEEE PES General Meeting, Providence, RI, USA, 25–29 July 2010; pp. 1–7.
- 42. Mohammed, N.N.; Abdulazeez, A.M. Evaluation of partitioning around medoids algorithm with various distances on microarray data. In Proceedings of the 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) (TEMS-ISIE) IEEE, Exeter, UK, 21–23 June 2017; pp. 1011–1016.
- Cao, D.; Yang, B. An improved k-medoids clustering algorithm. In Proceedings of the 2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE) IEEE, Singapore, 26–28 February 2010; pp. 132–135.
- Reynolds, A.P.; Richards, G.; Rayward, S.V.J. The Application of k-Medoids and Pam to the Clustering of Rules. In Proceedings of the 5th International Conference on Intelligent Data Engineering and Automated Learnin, Exeter, UK, 25–27 August 2004; pp. 173–178.
- 45. Robertson, S.E.; Jones, K.S. Relevance weighting of search terms. J. Am. Soc. Inf. Sci. 1976, 27, 129–146. [CrossRef]
- 46. Son, C.; Jeon, J.; Park, Y. Development of a GTM-based patent map for identifying patent vacuums. *Expert Syst. Appl.* **2012**, *39*, 2489–2500. [CrossRef]
- 47. Kayastha, S.; Horvath, D.; Gilberg, E. Privileged structural motif detection and analysis using generative topographic maps. *J. Chem. Inf. Model.* **2017**, *57*, 1218–1232. [CrossRef] [PubMed]
- 48. Sidorov, P.; Davioud-Charvet, E.; Marcou, G. AntiMalarial Mode of Action (AMMA) Database: Data Selection, Verification and Chemical Space Analysis. *Mol. Inform.* **2018**, *37*, 1800021. [CrossRef]
- 49. Kaneko, H. Sparse Generative Topographic Mapping for Both Data Visualization and Clustering. J. Chem. Inf. Model. 2018, 58, 2528–2535. [CrossRef]
- 50. Kim, H.; Hong, S.; Kwon, O.; Lee, C. Concentric diversification based on technological capabilities: Link analysis of products and technologies. *Technol. Forecast. Soc. Chang.* **2017**, *118*, 246–257. [CrossRef]
- 51. Owen, J.R.; Nabney, I.T.; Medina-Franco, J.L.; López-Vallejo, F. Visualization of molecular fingerprints. *J. Chem. Inf. Modeling* **2011**, *51*, 1552–1563. [CrossRef] [PubMed]
- 52. Billinton, R.; Chen, H.; Ghajar, R. Time-series models for reliability evaluation of power systems including wind energy. *Microelectron. Reliab.* **1996**, *36*, 1253–1261. [CrossRef]
- 53. Chen, P.; Pedersen, T.; Bak-Jensen, B.; Chen, Z. ARIMA-based time series model of stochastic wind power generation. *IEEE Trans. Power Syst.* 2009, 25, 667–676. [CrossRef]
- 54. Jeong, S.; Yoon, B. A systemic approach to exploring an essential patent linking standard and patent maps: Application of generative topographic mapping (GTM). *Eng. Manag. J.* **2013**, *25*, 48–57. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).