Article

Identifying Technology Opportunities for Electric Motors of Railway Vehicles with Patent Analysis

Yunkoo Cho [1], Young Jae Han [2,*], Jumi Hwang [3], Jiwon Yu [3], Sangbaek Kim [4], Chulung Lee [5,*], Sugil Lee [6] and Kyung Pyo Yi [6]

1   Adam International Patent & Law Office, 1313, 127 Beobwon-ro, Songpa-gu, Seoul 05836, Korea; ykcho@adampatent.com
2   Railroad Test & Certification Division, Korea Railroad Research Institute, 176 Cheoldobangmulgwan-ro, Uiwang-si 16105, Gyeonggi-do, Korea
3   Department of Industrial Management Engineering, Korea University, 145 Anam-ro, Seongbuk, Seoul 02841, Korea; hwangj0m@korea.ac.kr (J.H.); vermouth28@korea.ac.kr (J.Y.)
4   Department of Packaging, Yonsei University, 1, Yeonsedae-gil, Heungeop-myeon, Wonju-si 26493, Gangwon-do, Korea; ksba1019@yonsei.ac.kr
5   School of Industrial and Management Engineering, Korea University, 145 Anam-ro, Seongbuk, Seoul 02841, Korea
6   Smart Electrical & Signaling Division, Korea Railroad Research Institute, 176 Cheoldobangmulgwan-ro, Uiwang-si 16105, Gyeonggi-do, Korea; sglee@krri.re.kr (S.L.); kpyi82@krri.re.kr (K.P.Y.)
*   Correspondence: yjhan@krri.re.kr (Y.J.H.); leecu@korea.ac.kr (C.L.); Tel.:+82-31-460-5614 (Y.J.H.); +82-2-3290-3395 (C.L.)

Abstract: An electric motor is a device that changes electrical energy into mechanical energy for railway vehicles. When developing the electric motor, it used to be developed simply for structures or control methods of the motor itself without considering convergence with other devices or technologies. However, as the railway vehicles become more advanced, technology development through convergence with other devices or technologies is spreading. Therefore, based on patent data related to the electric motors applied to the railway vehicles, this research aims to carry out technical forecasting for establishing research and development (R and D) direction for new technologies by predicting vacant technologies from the point of view of technology convergence. In other words, we studied how to find the vacant technologies in a field of convergence technology for the electric motor of the railway vehicles by analyzing the patent data. More specifically, we search the patents data associated with the electric motor of the railway vehicle that contain multiple IPC codes, and use multiple IPC codes to determine the field of convergence technology. In addition, we extract keywords from the patents data related to each of the determined convergence technologies and define the vacant technologies by interpreting the field of convergence technology and the extracted keywords.

Keywords: vacant technology; convergence technology; electric motor; patent analysis

1. Introduction

Railway vehicles have advantages over other vehicles in that they can transport many passengers or goods quickly and can carry a long distance at a relatively low cost. The railway vehicles include a variety of devices or systems, of which electric motors are related to the operation of railway vehicles. The electric motors are very important device in the railway vehicles because they control acceleration and deceleration of the railway vehicles. Therefore, it is very important to predict the direction of technological development of the electric motors in order to predict the direction of technological development of the railway vehicles.

The electric motor areas of the railway vehicles comprise many complex devices or systems. As technology has developed and diversified, the importance of convergence

technologies among these devices or systems has increased. In addition, the importance of convergence technologies among the technologies of railway vehicles and technologies from other fields has also increased [1,2]. Therefore, predicting vacant technologies for convergence technologies related to the electric motors is very important to lead the technology areas related to the railway vehicles.

In the past, the vacant technologies were forecasted based on expert opinions using Delphi, scenario methods, etc. [3]. However, the forecasting of the vacant technologies by the experts may involve subjective judgement by the expert, so there is a problem that the conclusions can vary depending on who forecasts them. To overcome this problem, various studies continue to forecast the vacant technologies using patent data [4].

Therefore, in this study, we intend to derive the R and D strategy in an objective and quantitative method by forecasting the vacant technology in the field of the convergence technology related to the electric motor of the railway vehicles using the patent data. The patent data containing multiple IPC codes are defined as the patent data related to convergence technology. That is, we define the field of the convergence technology where the vacant technology can be derived using the patent data containing multiple IPC codes. We then derive the vacant technology for each convergence technology field using keywords extracted from the patent data corresponding to the convergence technology field. Using these derived vacant technologies, the R and D strategy was established by forecasting the vacant technologies for each field of convergence technology related to the electric motor of the railway vehicles. Therefore, this study will allow anyone to derive the vacant technologies in the field of the convergence technology and establish an R and D strategy without using the experts.

This paper is organized as follows: Section 2 reviews the case studies on patent analysis and patent map. Section 3 focuses on the overall research framework and specific methods for each step. Section 4 presents the results of analyzing the patent data, and Section 5 presents the conclusions of the results.

2. Literature Review

The electric motor areas of the railway vehicle can be defined as comprehensive engineering that requires many complex systems. Therefore, companies and research institutions have struggled to establish the direction of R and D because their survival in industry depends on the decision making. Therefore, previous studies attempted to explore appropriate new ideas that could contribute to the research direction [5,6]. However, there are limitations in that the studies have been conducted with qualitative approaches or a specific country aspect. The qualitative approaches can derive various interpretation for different researchers because the level of experience or knowledge vary depending on the researchers. Thus, results can be derived in various ways, depending on how the researcher determines. Studies as an aspect of specific country have limitations in that the technical trend cannot be reflected on the practical R and D planning because the research cannot cover the overall trend of global industry. Therefore, analysis from a global perspective based on a quantitative approach should be performed.

Patent documents contain textual information and bibliographic information [7]. The textual information (i.e., claims, description, etc.) have been employed to identify technological details of specific patent. In many previous studies, text mining has been attempted to extract significant keywords from patent documents, and a keyword vector have been generated based on the extracted keywords [4]. The vector can be substituted for the description of the characteristics of the patent. The bibliographic information (i.e., classifications, patent number, application number, etc.) have been employed to speculate technological trends based on statistical analysis [8]. Patents must be classified by international patent classification (IPC) code that represent the technology fields of each patent. If a patent is classified to various IPC codes, it is possible to interpret that there are some technological associations. A high number of patents classified by an IPC code have regarded high interest in the technology represented by this IPC code.

The concept of patent map has been employed to extract technical information and managerial implication by analyzing, classifying, and arranging the raw patent data. The patent maps relying on simple statistical tools have some limitations in explanatory capacity and operational efficiency due to the complexity of the data [9,10]. As machine learning-based approaches have been employed in analyzing patent data to develop patent map that can objectively detect patent vacuums [8]. The patent vacuums are considered vacant fields of technology with high potential to emerge in the near future.

Various researchers, therefore, applied machine learning techniques (i.e., principal component analysis (PCA), self-organizing map (SOM), and generative topographic mapping (GTM)) in developing the patent maps [11–13]. Machine learning-based patent maps can visualize the data space informatively by reducing the multi-dimensional data to two or three dimensions. The PCA-based patent map project the patent data on the two-dimensional space by detecting significant principal components based on linear combinations of the multi-variables [11,14]. The principal components can account for the most significant variations within the database. However, the PCA-based patent maps have some limitations in that they are difficult to automatically define the patent vacuums because there are no rules to define outliers [13].

SOM-based patent maps can locate patent information into discrete nodes by visualizing data as neurons from multi-dimensional neurons into two-dimensional neurons [15,16]. Visualization with color contrast were applied to show the similarities and differences between nodes. Therefore, researchers can detect the location of patent vacuums according to color scale in the SOM-based patent map. Since a great deal of time was required to extract significant and valuable idea for practical R and D planning, this process to detect patent vacuums have been regarded as limitations of the SOM-based patent map.

GTM-based patent map can project multi-dimensional data space into low-dimensional latent space and vice versa, and patent documents can also be located into the discrete distribution. More importantly, vectors for patent vacuums in the map can be estimated due to the inverse mapping algorithm based on Bayes' theorem [17]. Therefore, GTM-based patent map can overcome the aforementioned limitations because it can automatically detect patent vacuums and interpret them in an objective way. According to literatures, the GTM-based patent map is a suitable method for quantitatively generating patent maps in that it does not require manual investigation by researchers for interpretation of a patent vacuums [13].

Therefore, this study employed GTM algorithm to develop the patent map. As aforementioned, GTM-based patent map can automatically detect patent vacuum area. Despite some advantages of machine learning-based algorithm, the result has been inadequate for researchers to make decisions. In this regards, this study attempted to identify technical relationships in terms of technology convergence. Previous studies approached classification of patents. A patent can be characterized according to some IPC codes. Therefore, patent analysis based on the classification information has been conducted to predict future technological trends by analyzing the technological relationships based on social network analysis (SNA) and association rule mining (ARM).

The network according to SNA can be constructed using citation or classification information in the patent data. Frequently cited patents can be regarded as more important patents since the citation information can be qualitative indicator [18]. Patents that are frequently cited in subsequent patents are technically important because they imply that new technologies or related patents based on them may be filed. By analyzing the knowledge flow between the citing patent and the cited patent, it is possible to identify the relationship between technology or researchers [19].

However, the citation-based SNA has several inconvenience [20]. First, the citation network merely demonstrates individual connections between two particular patents, which interferes with understanding the overall relationship between all patents. Second, the network can be simply visualized, but there are difficulties in interpreting the internal relationships between patents or technical functions. Finally, the cited patents have an

average time-lag of 10 years from the citing patent [21]. Therefore, the latest technical trends cannot be reflected properly because the latest patents do not have enough time to be cited.

The IPC co-occurrence, two or more IPC codes simultaneously assigned to a patent document, have been considered that there is relationship with technical knowledge [22,23]. For instance, the IPCs that occur frequently can be interpreted as existing the strength of a relationship. Furthermore, there is no time-lag between citing and cited patents, since patent number and classification information are registered at the same time. ARM was then applied to identify relationships from patent data according to co-occurrences [24]. ARM mines frequent IPCs that they have meaningful relationships. ARM can derive pattern of technical information in terms of technology convergence characteristics. However, the classic association rules have some difficulties in that the rules cannot reflect the proper relations. In order to address this limitation, weighted association rule mining (WARM) tools are proposed [25].

This research aims to explore the technology convergence in terms of vacant technology for the electric motors of the railway vehicles. Firstly, we develop IPC-based patent map by employing GTM algorithm that can automatically detect patent vacuum areas. Secondly, technical relationships for technology convergence can be identified by applying WARM to derive associated patterns that are estimated to emerge in the future. Therefore, we can obtain macroscopic information for promising technology convergence (i.e., IPC-based results). In order to grasp specific technical information, a text mining tool is finally applied to extract technical keywords at a micro level. These keywords can be defined as core technologies with high potential. Consequently, this study suggest that technological competitiveness can be reinforced by proposing technological areas with high potential for development through the analyzed promising technologies and providing relevant information to relevant parties in establishing R and D strategies.

3. Methodology

Figure 1 shows the research framework for finding the vacant technologies in the field of electric motors of the railway vehicles. First, we collect patents in the field of electric motors from patent databases of USPTO. The IPC information of the collected patent documents is extracted. Second, the IPC vector is analyzed to create a GTM-based patent map and derive the patent vacuum that the technology has. The derived patent map analyzes the IPC of the patent vacuum by applying the inverse function. Third, analyze the patent vacuum through inverse mapping and applied WARM to interpret the relationship between technologies. Finally, we analyze the significance of the vacant technique derived by applying the text mining technique.

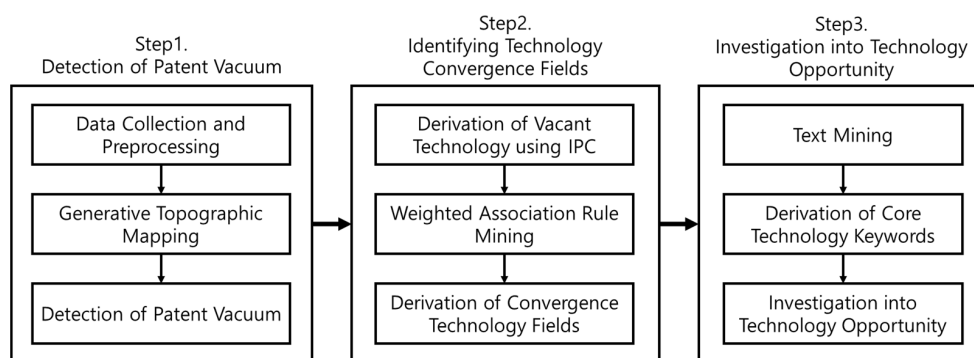**Figure 1.** Research framework.

3.1. Data Collection and Preprocessing

Many companies have patent databases in many countries to protect their rights regarding their technologies. The database with the most patents filed by many companies is

the USPTO [26]. In this paper, we extracted the patent data of the electric motor technology by applying for USPTO patents.

The extracted patent consists of title, abstract, IPC, image, etc. The IPC code indicates the technology group to which the patent belongs. The IPC code consists of sections, classes, subclasses, main groups, and subgroups, and through the information, you can find out the technical information of the patents. In this paper, the IPCs of patent documents are extracted and used to analyze the field of the convergence technologies.

3.2. Generative Topographic Mapping

GTM is an algorithm developed by Bishop et al. (1998), a nonlinear mapping methodology using the characteristics of data that map data to a lower dimension in multi-dimensional space [17]. GTM utilizes latent variables to model various correlations between many variables in the input data [27]. Each grid expressed as a latent variable displays data, and blank grids are automatically significant as empty spaces. It is a methodology that overcomes the limitations of PCA that require qualitative detection and definition of blank areas and SOM that need to be interpreted qualitatively for blank areas. In particular, GTM, which features nonlinearity, can identify the relationship between potential and observational variables through inverse mapping [13]. The basic principle of GTM is to dimension data in multi-dimensional space by projecting it in a low-dimensional potential grid of nodes of the size of $k \times k$. Shown in Figure 2, nodes are represented by radial basis functions (RBFs) in a grid of low-dimensional potentials, where RBFs included in the function $y(x; W)$ that connects the two spaces.
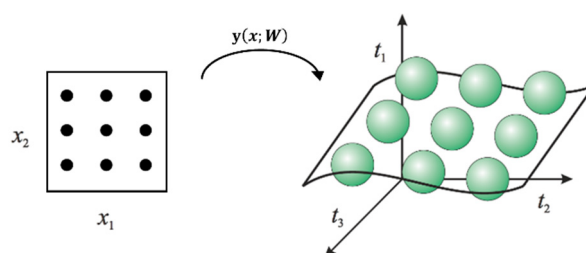

Figure 2. Basic concept of the GTM.

The parameters weight and noise are estimated through the EM algorithm. The EM algorithm consists of two stages, an E-step and an M-step. In the E-step, the initial values of the two parameters are used to calculate the responsibilities of the Gaussian mixed model. Reliability allows the calculation of the post-probability distribution between potential variables and data. In the M-step, the reliability function is updated to update the parameters. The error function is used as the objective function to update, where the updated direction reflects the characteristics of the data.

Finally, for each M mapped on GTM, the probability of M present in node K, that is, the probability matrix R (M, K) that provides the reflectivity of node K for molecule M, is calculated. This matrix can be used to identify the characteristics of each node and to visualize the GTM [28].

In this paper, we used IPC code extracted from patent data to generate the content-IPC Matrix and then applied GTM. Each IPC code represents the nature of the vector, and the resulting IPC vector has an N dimension. The data mapped to a potential two-dimensional space, and as a result, the nodes in the mapped map are joined together by patents on a similar topic based on IPC, enabling blank technical analysis through nodes on the map.

3.3. Weighted Association Rule Mining

The classic ARM has been considered one of the most useful mining tools that can discover patterns based on co-occurrence in the dataset [29]. A dataset contains numerous transactions that consist of a set of items. Association rule is an implication to find frequent item sets. In this research, we define a patent as a transaction, and the IPCs classifying a

patent can be defined as items included in the transaction. As aforementioned, WARM is applied to improve the limitations [30]. The weight is assigned to each IPC by calculating the number of patents including a specific IPC. Transaction weight (TW) is calculated by dividing the total weight of the items by the item counts included in the transaction. The weighted itemset (WI) is calculated by dividing the summed weight of TW according to specific transaction by the total weight of the transactions. Therefore, we can identify the most associated IPC combinations that could be candidates, when planning practical R and D strategies.

3.4. Text Mining

In this research, core keywords were derived using TF-IDF from the text mining method. TF-IDF is an algorithm for extracting the core word of each document by weighting each term based on the term frequency and the inverse document frequency. In other words, TF-IDF is a method of comparing the term frequency in each document with the term frequency in the entire document and applying important weighting to each word to determine the importance of words [31].

$$TF - IDF = tf_{d,t} \times log\left(\frac{N}{df_t}\right) \tag{1}$$

d: Document
t: Term
N: Total number of documents
$tf_{d,t}$: Frequency of words in a particular document
df_t: Frequency of words in the entire document

Each $tf_{d,t}$ represents the frequency at which each term (*t*) appears in the document concerning all words appearing in the specific document (*d*), and the higher the $tf_{d,t}$, the higher the importance in the document. df_t represents the frequency of the word appearing in the whole document (*N*), and the higher the df_t, the more frequently the word is used in all documents in general [32]. The *TF-IDF* algorithm is a method to reduce the importance of a word by determining that the word frequently appears in a particular document is not so important and determining df_t from the whole document (*N*), as shown in Equation (1) [33]. In other words, the *TF-IDF* applies weights to the words that appear frequently in all documents, and to the words that appear frequently in specific documents. Therefore, since the derived high-level words can be said to be the keywords for technologies that have not had many patent applications in each IPC technical field, the derived high-level words can be used as keywords that can define blank descriptions [34].

4. Results

4.1. Data Collection and Preprocessing

In this research, we selected keywords related to 'electric motor technology of railway vehicles' through expert advice. Based on selected keywords, patents related to 'electric motor technology of railway vehicles' were collected through the USPTO from 1990 to 2020. A total of 672 valid patents were selected through noise filtering and used for analysis. The IPC included in each patent was classified as subclass units (four-digit), resulting in a total of 119 IPCs. In this research, IPC codes are used to specify the field of technology, so instead of the entire IPC code, we utilize four-digit IPC codes that can specify the field of technology. We calculated the frequency for each IPC and only 15 IPCs with greater frequency than average were extracted to generate the patent-IPC matrix, as shown in Table 1.

Table 1. Patent-IPC matrix.

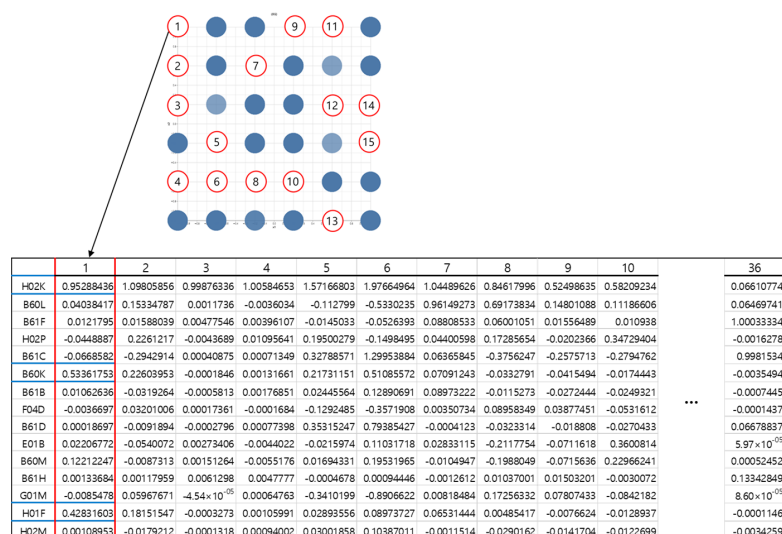|	H02K	B60L	B61F	H02P	B61C	B60K	. . .	H02M
Patent1	1	0	0	0	0	0		0
Patent2	1	1	0	0	0	0		0
Patent3	1	0	0	0	0	0		0
Patent4	0	0	0	0	0	0		0
Patent5	0	0	0	0	1	0		0
Patent6	0	0	0	0	1	0		0
Patent7	0	0	0	1	0	0		0
Patent8	1	0	0	0	0	0		0
. . .							. . .	
Patent672	0	0	0	0	0	0		0

4.2. Detection of Patent Vacuum

The GTM algorithm was applied to create a patent map and to derive a patent blank area. To create a GTM-based content map, you need to determine the appropriate map size (K). If the size of the map is too large, the sparse map is derived, resulting in too many blank areas, and if the size is too small, the blank area may not be derived [15]. In this research, the K value was set from 5 to 10 to determine the optimal K value. Shown in Figure 3, the optimal K set to 6, and the experiment resulted in a total of 15 blank nodes.
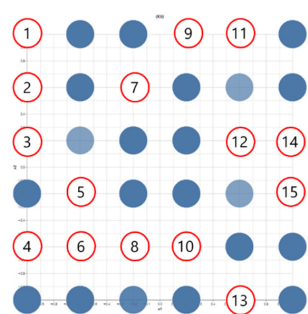
Figure 3. GTM-based patent map.

Through inverse mapping, GTM can express each patent space as a new vector space. The process of Inverse mapping is as shown in Figure 4. At this time, IPC with a probability value of 0.4 or higher for each IPC per node was determined to be the technical field associated with the blank technology of the patent.

	1	2	3	4	5	6	7	8	9	10		36
H02K	0.95288436	1.09805856	0.99876336	1.00584653	1.57166803	1.97664964	1.04489626	0.84617996	0.52498635	0.58209234		0.06610774
B60L	0.04038417	0.15334787	0.0011736	-0.0036034	-0.112799	-0.5330235	0.96149273	0.69173834	0.14801088	0.11186606		0.06469741
B61F	0.0121795	0.01588039	0.00477546	0.00396107	-0.0145033	-0.0526393	0.08808533	0.06001051	0.01556489	0.010938		1.00033334
H02P	-0.0448887	0.2261217	-0.0043689	0.01095641	0.19500279	-0.1498495	0.04400598	0.17285654	-0.0202366	0.34729404		-0.0016278
B61C	-0.0668582	-0.2942914	0.00040875	0.00071349	0.32788571	1.29953884	0.06365845	-0.3756247	-0.2575713	-0.2794762		0.9981534
B60K	0.53361753	0.22603953	-0.0001846	0.00131661	0.21731151	0.51085572	0.07091243	-0.0332791	-0.0415494	-0.0174443		-0.0035494
B61B	0.01062636	-0.0319264	-0.0005813	0.00176851	0.02445564	0.12890691	0.08973222	-0.0115273	-0.0272444	-0.0249321		-0.0007445
F04D	-0.0036697	0.03201006	0.00017361	-0.0001684	-0.1292485	-0.3571908	0.00350734	0.08958349	0.03877451	-0.0531612		-0.0001437
B61D	0.00018697	-0.0091894	-0.0002796	0.00077398	0.35315247	0.79385427	-0.0004123	-0.0323314	-0.018808	-0.0270433		0.06678837
E01B	0.02206772	-0.0540072	0.00273406	-0.0044022	-0.0215974	0.11031718	0.02833115	-0.2117754	-0.0711618	0.3600814	...	5.97×10^{-05}
B60M	0.12212247	-0.0087313	0.00151264	-0.0055176	0.01694331	0.19531965	-0.0104947	-0.1988049	-0.0715636	0.22966241		0.00052452
B61H	0.00133684	0.00117959	0.0061298	0.0047777	-0.0004678	0.00094446	-0.0012612	0.01037001	0.01503201	-0.0030072		0.13342849
G01M	-0.0085478	0.05967671	-4.54×10^{-05}	0.00064763	-0.3410199	-0.8906622	0.00818484	0.17256332	0.07807433	-0.0842182		8.60×10^{-05}
H01F	0.42831603	0.18151547	-0.0003273	0.00105991	0.02893556	0.08973727	0.06531444	0.00485417	-0.0076624	-0.0128937		-0.0001146
H02M	0.00108953	-0.0179212	-0.0001318	0.00094002	0.03001858	0.10387011	-0.0011514	-0.0290162	-0.0141704	-0.0122699		-0.0034259

Figure 4. Inverse mapping process of GTM.

The IPC for each patent space derived from the inverse mapping result is as shown in Table 2.

Table 2. IPC Results.

Patent Vacuum	IPC
1	H02K, B60K, H01F
2	H02K
3	H02K
4	H02K
5	H02K
6	H02K, B61C
7	B60L
8	Blank
9	B60L
10	B61D
11	B60L
12	B60L, H02P, B61C
13	B61F, B61C, B60K, B61D
14	B60L, B61C
15	B61F

4.3. Identifying Technology Convergence Field

Shown in Table 3, the technical field of blank technology derived from the GTM-based patent map can be defined as B60K, B60L, B61C, B61D, B61F, H01F, H02K, H02P, based on which WARM is applied, and the weighted support value is 0.006 or higher.

Table 3. WARM Result.

	Itemset	Item Number	Summed Weight of Items	Transaction Weight	Weighted Support
1	H02K, B60L	2	0.391	0.195	0.007
2	B61C, B60K	2	0.377	0.189	0.007
3	B61C, H02K, B60L	3	0.550	0.183	0.007
4	B61F, H02K, B60L	3	0.512	0.171	0.006
5	B61F, B60L	2	0.340	0.170	0.006
6	B61F, B61C, B60L	3	0.499	0.166	0.006
7	B61C, H02K	2	0.331	0.166	0.006

4.4. Investigation into Technology Opportunity

Reclassified patent data by Itemset derived by Weighted Association Rule Mining. Itemset1 selects patent data containing H02K and B60L by IPC code, Itemset2 selects patent data containing B61C and B60K by IPC code. Itemset3 selects patent data containing B61C, H02K, and B60L by IPC code, Itemset4 selects patent data containing B61F, H02K and B60L by IPC code. Itemset5 selects patent data containing B61F and B60L by IPC code, Itemset6 selects patent data containing B61F, B61C, and B60L by IPC code, Itemset7 selects patent data containing B61C and H02K by IPC code.

Text Mining was also applied based on the summary and claim information of the patent data sorted by Itemset to derive detailed technology keywords. Shown in Table 4, top 10 keywords by item derived as a result of text mining are as follows.

Table 4. Top 10 keywords of each itemset.

Itemset 1	Itemset 2	Itemset 3	Itemset 4	Itemset 5	Itemset 6	Itemset 7
filter	Rotor	speed	vehicle	alternation	levitation	damper
transducer	Highspeed	supercapacitor	lightweight	aluminum	track	decompression
suspension	Hollow	inverter	synchronous	carrier	induction	drawing
valve	Shaft	diesel	rail	construction	frame	motor
flashover	Cooling	asynchronous	reduction	frame	bolster	spring
coil	Winding	reducer	schematic	gate	arc	steep
chain	Yoke	wiring	dieselelectric	intervals	passive	synchronizer
multivoltage	Linear	magnet	copper	legs	connecting	track
sensor	Magnet	permanent	surface	light	vertical	trolley
monitor	Supply	small	aluminum	motor	running	wheel

4.5. Results

The IPC code associated with vacant techniques derived from the GTM-based patent map are defined, as shown in Table A1 of Appendix A. Multiple IPC codes used to analyze the field of technology for each Itemset. Using keywords extracted by itemset, the vacant technology in the defined technology field was derived. In the process of deriving the vacant technology, the vacant technology was defined except for unrelated keywords.

Itemset 1 is a field of convergence technology related to H02K, B60L, and the top 10 keywords extracted are filter, transducer, suspension, valve, flashover, coil, chain, multi-voltage, sensor, monitor. Of the IPC codes associated with Itemset 1, H02K is a technology for rotary-electrical machinery including electric motors, and B06L is a technology for electrical propulsion, power supply, charging, control, safety devices, and electrical braking in vehicles. Therefore, the technology related to Itemset 1 can be seen as a convergence technology for electrical propulsion, power supply, charging, control, safety devices, and electrical braking using the electric motors of the railway vehicles. Based on the keywords extracted concerning Itemset 1, the vacant technology can be seen as a monitoring technology such as filters, transducers, valves, coils, etc., using sensors. Therefore, the vacant technology derived from the IPC codes and keywords related to Item 1 is the monitoring technology such as filters, transducers, valves, coils, etc., using sensors in the field of technology related to electrical propulsion, power supply, charging, control, safety devices, and electrical braking using the electric motors of the railway vehicles.

Itemset 2 is a field of technology related to B61C, B60K, and the top 10 keywords extracted are rotor, highspeed, hollow, shaft, cooling, winding, yoke, linear, magnet, supply. IPC code B61C is a technology for power devices applied to railway vehicles and B60K is a technology for the deployment or control of power transfer devices in vehicles. Therefore, the technology related to Itemset 2 can be seen as a convergence of power devices applied to railway vehicles, etc., and power transmission devices transmitting power from power units. Based on the keywords extracted concerning Itemset 2, the vacant technology can be seen as a high-speed cooling technology of permanent magnet motors. Therefore, the vacant technology derived from the IPC codes and keywords related to Item 2 is the high-speed cooling technology of permanent magnet motors in the field of technology related to the power devices or the power transmission devices of the railway vehicles.

Itemset 3 is a field of technology related to B61C, H02K, and B60L. The top 10 keywords extracted are speed, supercapacitor, inverter, diesel, asynchronous, reducer, wiring, magnet, permanent, small. IB61C is a technology for power devices applied to the railway vehicles. H02K is a technology for rotary-electric machinery including the electric motors. B06L is a technology for electric propulsion, power supply, charging, control, safety devices, electrical braking in vehicles including railway vehicles. Therefore, the technology related to Itemset 3 can be seen as a convergence technology for electrical propulsion, power supply, charging, control, safety devices, and electrical braking using the electric motors of the railway vehicles. Based on the keywords extracted concerning Itemset 3, the vacant technology can be seen as a speed control technology of the permanent magnet motors

and a minimization technology of the permanent magnet motors. Therefore, the vacant technology derived from the IPC codes and keywords related to Itemset 3 is the speed control technology of the permanent magnet motors and a minimization technology of the permanent magnet motors in the field of technology related to electrical propulsion, power supply, charging, control, safety devices, and electrical braking using the electric motors of the railway vehicles.

Itemset 4 is a field of technology related to B61F, H02K, and B60L. The top 10 keywords extracted are vehicle, lightweight, synchronous, rail, reduction, schematic, dieselelectric, copper, surface, aluminum. B61F is a technique for the suspensions, vehicle protection, and obstacle removal of the railway vehicles. H02K is a technology for rotary-electric machinery are including the electric motors, and B06L is a technology for electrical propulsion, power supply, charging, control, safety devices, and electrical braking in the railway vehicles. Therefore, the field of technology related to Itemset 4 can be seen as a convergence of the technology related to the suspension, vehicle protection, and obstacle removal of the railway vehicles and the control technology of the electric motors. Based on the keywords extracted concerning Itemset 4, the vacant technology can be seen as a lightening technology using the alternative materials, such as aluminum, for railway vehicles. Therefore, the vacant technology derived from the IPC codes and keywords related to Itemset 4 is the technology to lighten the suspensions, vehicle protection, and obstacle removal of the railway vehicles using the alternative materials, such as aluminum, to improve the performance of electrical propulsion, power supply, charging, control, safety devices, and electrical braking in the railway vehicles.

Itemset 5 is a field of technology related to B61F and B60L. The top 10 keywords extracted are alternation, aluminum, carrier, construction, frame, gate, intervals, legs, light, motor. B61F is a technology on the suspensions, vehicle protection, and obstacle removal of the railway vehicles. B06L is a technology for electric propulsion, power supply, charging, control, safety devices, electrical braking in the railway vehicles. Therefore, the field of technology related to Itemset 6 can be seen as a convergence of the technology related to the suspension, vehicle protection, and obstacle removal of the railway vehicles and the control technology of the electric motors. Based on the keywords extracted concerning Itemset 5, the vacant technology can be seen as a lightening technology using the alternative materials, such as aluminum, for railway vehicles. Therefore, the vacant technology derived from the IPC codes and keywords related to Itemset 5 is the technology to lighten the suspensions, vehicle protection, and obstacle removal of the railway vehicles using the alternative materials, such as aluminum, to improve the performance of electrical propulsion, power supply, charging, control, safety devices, and electrical braking in the railway vehicles.

Itemset 6 is a field of technology related to B61F, B61C, and B60L. The top 10 keywords extracted are levitation, track, induction, frame, bolster, arc, passive, connecting, vertical, running. B61F is a technique for the suspensions, vehicle protection, and obstacle removal of the railway vehicles. B61C is the technology for power devices applied to the railway vehicles, and B06L is the technology for electric propulsion, power supply, charging, control, safety devices, and electrical braking in the railway vehicles. The fields of technology analyzed through B61F, B61C, and B60 are analyzed as a convergence of technologies related to the suspension of the railway vehicles and the control technology of power devices applied to the railway vehicles. Based on the keywords extracted concerning Itemset 6, the technology of magnetic levitation train with power unit or the technology related to the surface or frame of the railway vehicle is analyzed as the vacant technology. Therefore, the vacant technology derived from the IPC codes and keywords related to Itemset 6 is the control technology of the magnetic levitation train with the electric motor or a technology related to the surface or the frame of the magnetic levitation train.

Itemset 7 is a field of technology related to B61C, H02K. The top 10 keywords extracted are damper, decompression, drawing, motor, spring, steep, synchronizer, track, trolley, wheel. B61C is a technology for power devices applied to the railway vehicles, and H02K is a technology for rotary-electric machinery including the electric motors. H02K contains the

electric motor used in various devices, and B61C includes a power unit used in the railway vehicles. The technical fields commonly applied to H02K and B61C analyzed as to the electric motors fitted to other devices, including the railway vehicles. Therefore, Itemset 7 has analyzed a single technology related to electric motor, not a technology that combines motor-related technology with technology from other fields. The results were excluded without analysis of the keywords extracted because they did not conform to the subject matter of this research.

5. Conclusions

This research explored the patent data and looked for ways to derive the vacant technology in the field of convergence technology for the electric motor of the railway vehicles. Unlike other studies that analyzed vacant technologies alone, we derived vacant technologies in terms of convergence to derive future promising technologies that are further advanced. By applying the GTM methodology and WARM, we extract IPC codes and define the field of convergence technology us the extracted IPC codes. In addition, we extracted keywords from the patents related to each of the determined convergence technologies by using text mining. Lastly, we define the vacant technologies by interpreting the field of convergence technology and the extracted keywords.

Promising technologies defined through patent analysis are technologies that apply elastic materials to lighten/smaller vehicle frames and control objects and speed through monitoring. It also includes technology to applies permanent magnet motors, technology to transfer power at high speeds, and cooling technology. These promising technologies show that technologies related to small size, lightweight, permanent magnet, cooling, monitoring, low loss materials, and individual control can be core technology keywords in the future.

This paper confirmed that the miniaturization and weight reduction of electric motors is a global trend. It is analyzed that it is appropriate to adopt a permanent magnetic motor rather than an inductive motor for compact/lightweight motor. In order to secure the safety and reliability of the vehicle, it is analyzed that R and D is also necessary to establish a system for real-time condition monitoring of electric motors. The field related to this technology is a field with a high potential for development as the vacant technologies. If you apply the analyzed results to derive new technologies, you can gain a competitive advantage in related fields. This study analyzes promising technologies in the future, provides information on those technologies, provides information to help in R and D strategy establishment, and provides insights to researchers, strategy makers, and decision-makers. It can help you invent new technologies and plan your skills development.

However, the limitation of this research is that even in the field of technology that includes multiple IPC codes, if it contains different IPC codes related to it, it must be considered as a single technology, not a convergence technology. Nevertheless, this research is meaningful in that instead of relying on subjective judgments of experts to the vacant technology, IPC codes, and keywords extracted from patent data can be used to find the vacant technology in the field of convergence technology.

Appendix A

Table A1. Vacant technology by IPC code.

IPC Code	Definition
B60K	ARRANGEMENT OR MOUNTING OF PROPULSION UNITS OR OF TRANSMISSIONS IN VEHICLES; ARRANGEMENT OR MOUNTING OF PLURAL DIVERSE PRIME-MOVERS IN VEHICLES; AUXILIARY DRIVES FOR VEHICLES; INSTRUMENTATION OR DASHBOARDS FOR VEHICLES; ARRANGEMENTS IN CONNECTION WITH COOLING, AIR INTAKE, GAS EXHAUST OR FUEL SUPPLY OF PROPULSION UNITS IN VEHICLES
B60L	PROPULSION OF ELECTRICALLY-PROPELLED VEHICLES (arrangements or mounting of electrical propulsion units or of plural diverse prime-movers for mutual or common propulsion in vehicles B60K 1/00, B60K 6/20; arrangements or mounting of electrical gearing in vehicles B60K 17/12, B60K 17/14; preventing wheel slip by reducing power in rail vehicles B61C 15/08; dynamo-electric machines H02K; control or regulation of electric motors H02P); SUPPLYING ELECTRIC POWER FOR AUXILIARY EQUIPMENT OF ELECTRICALLY-PROPELLED VEHICLES (electric coupling devices combined with mechanical couplings of vehicles B60D 1/64; electric heating for vehicles B60H 1/00); ELECTRODYNAMIC BRAKE SYSTEMS FOR VEHICLES IN GENERAL (control or regulation of electric motors H02P); MAGNETIC SUSPENSION OR LEVITATION FOR VEHICLES; MONITORING OPERATING VARIABLES OF ELECTRICALLY-PROPELLED VEHICLES; ELECTRIC SAFETY DEVICES FOR ELECTRICALLY-PROPELLED VEHICLES
B61C	LOCOMOTIVES; MOTOR RAILCARS (vehicles in general B60; frames or bogies B61F; special railroad equipment for locomotives B61J, B61K)
B61F	RAIL VEHICLE SUSPENSIONS, e.g., UNDERFRAMES, BOGIES OR ARRANGEMENTS OF WHEEL AXLES; RAIL VEHICLES FOR USE ON TRACKS OF DIFFERENT WIDTH; PREVENTING DERAILING OF RAIL VEHICLES; WHEEL GUARDS, OBSTRUCTION REMOVERS OR THE LIKE FOR RAIL VEHICLES (for vehicles in general B60; axles or wheels B60B; vehicle tyres B60C)
H02K	DYNAMO-ELECTRIC MACHINES (dynamo-electric relays H01H 53/00; conversion of DC or AC input power into surge output power H02M 9/00)

References

1. Lee, J.; Won, J.; Kim, Y. A Study on the Construction of the Intergrated Platforms for the Railway IT Convergence. In *Proceedings of the Korean Society for Noise and Vibration Engineering Conference*; The Korean Society for Noise and Vibration Engineering: Seoul, Korea, 2010; pp. 455–456.
2. Cho, J. Study on Convergence Technique through Structural Analysis on the Axle of Railway Vehicle. *J. Korea Converg. Soc.* **2015**, *6*, 93–101. [CrossRef]
3. Kim, G.; Bae, J. A novel approach to forecast promising technology through patent analysis. *Technol. Forecast. Soc. Chang.* **2017**, *117*, 228–237. [CrossRef]
4. Lee, S.; Yoon, B.; Lee, C.; Park, J. Business planning based on technological capabilities: Patent analysis for technology-driven roadmapping. *Technol. Forecast. Soc. Chang.* **2009**, *76*, 769–786. [CrossRef]
5. Ko, N.; Yoon, J.; Seo, W. Analyzing interdisciplinarity of technology fusion using knowledge flows of patents. *Expert Syst. Appl.* **2014**, *41*, 1955–1963. [CrossRef]
6. Finger, M. Governance of competition and performance in European railways: An analysis of five cases. *Util. Policy* **2014**, *31*, 278–288. [CrossRef]
7. 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]
8. Jeong, Y.; Yoon, B. Development of patent roadmap based on technology roadmap by analyzing patterns of patent development. *Technovation* **2015**, *39*, 37–52. [CrossRef]
9. Park, I.; Jeong, Y.; Yoon, B.; Mortara, L. Exploring potential R&D collaboration partners through patent analysis based on bibliographic coupling and latent semantic analysis. *Technol. Anal. Strateg. Manag.* **2015**, *27*, 759–781.

10. Yoon, B.U.; Yoon, C.B.; Park, Y.T. On the development and application of a self–organizing feature map–based patent map. *R D Manag.* **2002**, *32*, 291–300. [CrossRef]
11. Jun, S.; Park, S.S.; Jang, D.S. Technology forecasting using matrix map and patent clustering. *Ind. Manag. Data Syst.* **2012**, *112*, 786–807. [CrossRef]
12. Lee, S.; Yoon, B.; Park, Y. An approach to discovering new technology opportunities: Keyword-based patent map approach. *Technovation* **2009**, *29*, 481–497. [CrossRef]
13. Wu, J.L.; Chang, P.C.; Tsao, C.C.; Fan, C.Y. A patent quality analysis and classification system using self-organizing maps with support vector machine. *Appl. Soft Comput.* **2016**, *41*, 305–316. [CrossRef]
14. 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]
15. Johnson, R.A.; Wichern, D.W. *Applied Multivariate Statistical Analysis*, 6th ed.; Prentice Hall: Upper Saddle River, NJ, USA, 1988.
16. Kohonen, T. The self-organizing map. *Proc. IEEE* **1990**, *78*, 1464–1480. [CrossRef]
17. Zhu, D. Bibliometric analysis of patent infringement retrieval model based on self-organizing map neural network algorithm. *Libr. Hi Tech* **2019**, *38*, 479–491. [CrossRef]
18. Bishop, C.M.; Svensén, M.; Williams, C.K. GTM: The generative topographic mapping. *Neural Comput.* **1998**, *10*, 215–234. [CrossRef]
19. 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]
20. Park, I.; Yoon, B. Technological opportunity discovery for technological convergence based on the prediction of technology knowledge flow in a citation network. *J. Informetr.* **2018**, *12*, 1199–1222. [CrossRef] [PubMed]
21. Yoon, B.; Park, Y. A text-mining-based patent network: Analytical tool for high-technology trend. *J. High Technol. Manag. Res.* **2004**, *15*, 37–50. [CrossRef]
22. Hall, B.H.; Adam, B. *The NBER Patent-Citations Data File: Lessons, Insights, and Methodological Tools. Patents, Citations, and Innovations: A Window on the Knowledge Economy*; NBER Working Paper; National Bureau of Economic Research: Cambridge, MA, USA, 2002; Volume 403. [CrossRef]
23. Tijssen, R.J. A quantitative assessment of interdisciplinary structures in science and technology: Co-classification analysis of energy research. *Res. Policy* **1992**, *21*, 27–44. [CrossRef]
24. Geum, Y.; Kim, C.; Lee, S.; Kim, M.S. Technological convergence of IT and BT: Evidence from patent analysis. *ETRI J.* **2012**, *34*, 439–449. [CrossRef]
25. Jun, S. IPC code analysis of patent documents using association rules and maps–patent analysis of database technology. In *Database Theory and Application, Bio-Science and Bio-Technology*; Springer: Berlin/Heidelberg, Germany, 2011; Volume 258, pp. 21–30.
26. Altuntas, S.; Dereli, T.; Kusiak, A. Analysis of patent documents with weighted association rules. *Technol. Forecast. Soc. Chang.* **2015**, *92*, 249–262. [CrossRef]
27. Engelsman, E.C.; van Raan, A.F. A patent-based cartography of technology. *Res. Policy* **1994**, *23*, 1–26. [CrossRef]
28. 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.
29. Agrawal, R.; Imieliński, T.; Swami, A. Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, USA, 26–28 May 1993; pp. 207–216.
30. Tao, F.; Murtagh, F.; Farid, M. Weighted association rule mining using weighted support and significance framework. In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 24–27 August 2003; pp. 661–666.
31. Aizawa, A. An information-theoretic perspective of tf–idf measures. *Inf. Process. Manag.* **2003**, *39*, 45–65. [CrossRef]
32. Juan, R. Using TF-IDF to determine word relevance in document queries. In Proceedings of the First Instructional Conference on Machine Learning, Washington, DC, USA, 21–24 August 2003; Volume 242, pp. 133–142.
33. Wu, H.C.; Luk, R.W.P.; Wong, K.F.; Kwok, K.L. Interpreting TF-IDF term weights as making relevance decisions. *ACM Trans. Inf. Syst.* **2008**, *23*, 1–37. [CrossRef]
34. Wen, Z. A comparative study of TF*IDF, LSI and multi-words for text classification. *Expert Syst. Appl.* **2011**, *38*, 2758–2765.