sustainability

MDPI

Article

Patent Keyword Analysis of Disaster Artificial Intelligence Using Bayesian Network Modeling and Factor Analysis

Sangsung Park and Sunghae Jun *

Department of Big Data and Statistics, Cheongju University, Chungbuk 28503, Korea; hanyul@cju.ac.kr
* Correspondence: shjun@cju.ac.kr; Tel.: +82-10-7745-5677

Received: 27 November 2019; Accepted: 7 January 2020; Published: 9 January 2020

Abstract: At present, artificial intelligence (AI) contributes to most technological fields. AI has also been introduced in the disaster area to replace humans and contribute to the prevention of disasters and the minimization of damages. So, it is necessary to analyze disaster AI in order to effectively make use of it. In this paper, we analyze the patent documents related to disaster AI technology. We propose Bayesian network modeling and factor analysis for the technology analysis of disaster AI. This is based on probability distribution and graph theory. It is also a statistical model that depends on multivariate data analysis. In order to show how the proposed model can be applied to a real problem, we carried out a case study to collect and analyze the patent data related to disaster AI.

Keywords: Bayesian statistics; disaster artificial intelligence; technology analysis; factor analysis; patent keyword analysis

1. Introduction

Technology analysis has played an important role in the management of technology (MOT) [1]. For example, an analysis of prior technology in target technological field is useful for research and development (R&D) planning in a company. To date, most companies have conducted various technology analyses for their sustainability. The convenient technology analysis has been performed based on the knowledge of domain experts. This approach is relatively subjective. This is because the results of the technology analysis may vary according to the composition of the expert group participating in the technology analysis of the specific domain [2,3]. In contrast, the technology analysis using machine learning algorithms or statistical analysis methods is relatively objective because it does not depend on the subjective experience of the expert group. Choi et al. (2016) proposed a hierarchical diagram of technology using regression modeling. They applied their proposed model to manage the sustainable technology [2]. Park and Jun (2017) studied a method of statistical technology analysis using social network analysis and time series clustering [3]. They carried out a case study to apply their method to sustainable technology analysis in three-dimensional (3D) printing technology.

In this paper, we also propose a technology analysis method using patent keyword analysis based on machine learning and statistics. The proposed method uses the Bayesian network model and factor analysis. Bayesian networks model is a popular graph model in machine learning and this model consists of nodes and arrows [4]. The nodes and arrows represent a random variable and probabilistic dependence between the nodes, respectively [5–7]. Factor analysis is a statistical method that reduces the dimensions of data [8,9]. This method uses the covariance structure between variables to extract factors representing variables and group each variable into corresponding factors [8,9]. In this paper, we built a patent analysis method using Bayesian network modeling and factor analysis for technology analysis. In addition, we considered disaster artificial intelligence (DAI) as a target technology domain

for our research. In our research, we defined the DAI as an AI to replace humans and contribute to the prevention of disasters and the minimization of damage. We also propose a statistical method for DAI technology. So, we collect the patent documents related to DAI for technology analysis using the proposed method.

The remainder of this paper is organized as follows. In Section 2, we show the patent technology analysis related to our study. We show the proposed method using Bayesian network modeling and factor analysis for technology analysis in Section 3. The following section provides the result of our case study for DAI technology analysis. In the conclusion, we conclude our research and describe our future works related to technology analysis.

2. Patent Technology Analysis

The patent contains a diverse and complete set of information regarding developed technology because the patent system protects the exclusive right of inventors to register their technology around the world [10]. Most research developers try to protect their technology by filling their technology with patent offices around the world. A patent document has a huge amount of technology information, such as the date of application, the inventor, the title, and the abstract of developed technology, citation, international patent classification (IPC) codes, claims, figures, and drawings, etc. Among them, much research on patent analysis used keywords and IPC data codes [11–14]. For example, they extracted the keywords from the collected patent documents related to target technology using the preprocessing techniques based on text mining. Using the extracted keywords, they constructed a matrix with patents (rows) and keywords (columns), and each element of the matrix was an occurred frequency of a patent keyword in each patent. This matrix is used as structured data for patent analysis using machine learning algorithms and statistical methods. In this paper, we also make the matrix from the retrieved patent documents related to DAI.

Since most of the information contained in patent documents is in text form, text mining is a very important analysis tool in patent data analysis [15]. Kim et al. (2019) studied text mining for patent data analysis, and they forecasted the emerging technologies in wireless power transfer. In this paper, we also proposed a framework for patent data analysis, including topic extraction, latent semantic analysis, clustering, and time series analysis [15]. Among the methods in the patent data analysis, the latent semantic analysis is a latent variable modeling used to find the association between unobserved variables using observed data [16]. This approach is effective for technology analysis using patent documents, because we can use the observed data based on the keywords in the patent document to find potentially embedded technologies. In this paper, we use factor analysis as a similar approach to latent semantic analysis. The factor analysis is also used to model the underlying structure in variables using multivariate normal distribution [17]. This generates the latent variables from the observed data. In our research, the latent variables are used to represent detailed technologies.

3. New Technology Analysis Using Bayesian Network Modeling and Factor Analysis

In this paper, we propose a new method for technology analysis using machine learning and statistics. To build the proposed technology analysis model, we combined Bayesian network modeling with factor analysis. The Bayesian networks model is one of the directed graphical models [9]. This model is essentially based on the following chain rule [8]:

$$p\left(X_1, X_2, \ldots, X_p\right) = p\left(X_p \middle| X_{p-1}, \ldots, X_1\right) p\left(X_{p-1} \middle| X_{p-2}, \ldots, X_1\right) \cdots p(X_1) \quad (1)$$

where $X_1, X_2, \ldots, X_p$ are random variables. This chain rule represents a joint discrete probability distribution. The random variable of this distribution has the discrete values with $(1, 2, \ldots, m)$ and

m is the maximum value that a random variable can have. Also, the Bayesian networks model is dependent on the following graph theory [6]:

$$G = (V, E) \tag{2}$$

Graph G consists of nodes, such as $V = \left(x_1, x_2, \ldots, x_p\right)$, and edges E. The V represents the p random variables. The E shows the probabilistic connections between two nodes (random variables). Consequently, the Bayesian networks model is called a directed acyclic graph (DAG), spanning from x_{i-1} to x_i. In our research, the node V represents the patent keywords extracted from the DAI patent documents as follows:

$$V = \left(Keyword_1, Keyword_2, \ldots, Keyword_p\right) \tag{3}$$

That is, we consider that the keywords are random variables $(x_i = Keyword_i)$ in the proposed model. We also compute the marginal distribution of $Keyword_1$ as follows:

$$p(Keyword_1) = \sum_{Keyword_2} \cdots \sum_{Keyword_p} p\left(Keyword_1, Keyword_2, \ldots, Keyword_p\right) \tag{4}$$

In the proposed model, we assume that all random variables (keywords) are mutually independent. In Bayesian networks, all random variables (keywords) are considered mutually independent for using the product rule of Equation (5). Through this, the Bayesian network finds the relationship between random variables. So, we show the joint distribution following the product distribution of each keyword.

$$p\left(Keyword_1, Keyword_2, \ldots, Keyword_p\right) = \prod_{i=1}^{p} p(Keyword_i) \tag{5}$$

So, we get the following marginal distribution of $Keyword_1$:

$$p(Keyword_1) = \left(\sum_{Keyword_p} p\left(Keyword_p\right) \cdots \sum_{Keyword_2} p(Keyword_2)\right) p(Keyword_1) \tag{6}$$

In Equation (6), each summation from $Keyword_p$ to $Keyword_1$ is computed independently. Therefore, we can factorize the joint distribution of $\left(Keyword_1, Keyword_2, \ldots, Keyword_p\right)$ as follows:

$$p\left(Keyword_1, Keyword_2, \ldots, Keyword_p\right) = \prod_{i=1}^{p} p\left(Keyword_i \middle| Keyword_1, \ldots, Keyword_{i-1}\right) \tag{7}$$

Using the factorization of the joint probability distribution in Equation (7), we consider three fundamental connections in Figure 1.

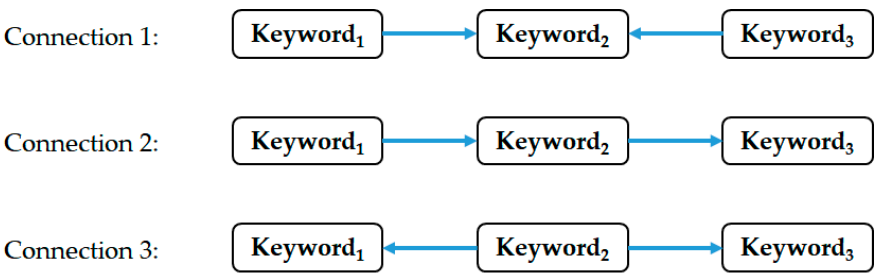

Figure 1. Three connection types of conditional dependences.

In this paper, we consider three connection types of conditional dependences between DAI keywords. The connection 1 of Figure 1 is a converging connection, which is defined as follows:

$$p(Keyword_1, Keyword_2, Keyword_3) = p(Keyword_2 \big| Keyword_1, Keyword_3) p(Keyword_1) p(Keyword_3) \quad (8)$$

In this connection, we know that the $Keyword_2$ depends on the joint distribution of $Keyword_1$ and $Keyword_3$. This means that $Keyword_1$ and $Keyword_3$ are not conditionally independent given $Keyword_2$. On the contrary, the $Keyword_1$ and $Keyword_3$ are independent given $Keyword_2$ in the connections 2 and 3. Connection 2 is called a serial connection and is represented as follows:

$$p(Keyword_1, Keyword_2, Keyword_3) = p(Keyword_3 \big| Keyword_2) p(Keyword_2 \big| Keyword_1) p(Keyword_1) \quad (9)$$

In addition, the conditional dependence of connection 3 is defined as follows:

$$p(Keyword_1, Keyword_2, Keyword_3) = p(Keyword_1 \big| Keyword_2) p(Keyword_3 \big| Keyword_2) p(Keyword_2) \quad (10)$$

This connection is called a diverging connection. These three connections are applied to our patent keyword analysis. The dependency and independency play an important role in the Bayesian visualization of this paper. For example, if DAG is $Keyword_A \rightarrow Keyword_B$ in Bayesian networks, $Keyword_A$ and $Keyword_B$ are called parent and child, respectively. In addition, the dependency of $Keyword_A$ and $Keyword_B$ is represented by the conditional probability $p(Keyword_B \big| Keyword_A)$. If $p(Keyword_B \big| Keyword_A)$ is equal to $p(Keyword_B)$, then $Keyword_A$ and $Keyword_B$ are independent of each other. Otherwise, they are dependent. So, three types of different connections, which are converging, serial, and diverging, are determined by the dependency and independency structure of the keywords (random variables). Next, we should construct the structured data suitable for machine learning algorithms and statistical methods for patent data analysis. In this paper, we built a matrix of patent documents and keywords. The rows and columns of this matrix are patents and keywords, respectively. The element of this matrix represents the frequency of specific keywords appearing in a particular patent document. This matrix is the structured data for our proposed method. In general, since the dimension of a patent-keyword matrix is very large, it is difficult to analyze it as it is. In particular, since the Bayesian network model represents the probabilistic connection state between keywords, when the number of keywords becomes too large, it is difficult to identify the probabilistic dependencies between the technological keywords. In order to solve this problem, this paper performs factor analysis on the structured data before performing Bayesian network modeling. Factor analysis reduces the dimensions of given data and enables efficient Bayesian network modeling. Factor analysis is defined as follows under the assumption of $m < p$ ($x \in R^p, f \in R^m$) [8,17]:

$$x - \mu = Af + \varepsilon \quad (11)$$

where x is the original variable and f is the factor (latent variables) representing underlying (latent) variables. Additionally, A and ε are factor loadings and error terms, respectively. Factor loading represents the distance (or similarity) between the original variable and the latent variable. The larger this value, the greater the similarity between the original variable and the corresponding factor. Thus, using factor loading results, p original variables can be grouped into m factors smaller than p. In this paper, we reduced the dimension of the patent-keyword matrix using this factor analysis. We define a factor analysis model for the dimension reduction of patent keyword data as follows:

$$Keyword_i - \mu_i = \sum_{j=1}^{m} a_{ij} f_j + \varepsilon_i, \; i = 1, 2, \ldots, p \quad (12)$$

where μ_i is the mean of $Keyword_i$, and a_{ij} is factor loading value between $Keyword_i$ and $Factor_j$ $\left(f_j\right)$. Therefore, we use factor analysis to reduce the dimension of the patent-keyword matrix and use this result to perform Bayesian network modeling. Finally, we present the proposed method step-by-step as follows.

Step 1 Collecting DAI patent documents from patent databases
Step 2 Preprocessing collected patent documents
Step 3 Constructing a patent–keyword matrix
Step 4 Performing factor analysis and Bayesian networks
Step 5 Finding a technological structure for DAI technology forecasting

In Step 1, we collect the patent documents related to DAI using a keyword searching equation. In this paper, we use the WIPS corporation (WIPSON) and the United States Patent and Trademark Office (USPTO) as patent databases around the world [18,19]. The keyword searching equation is shown in the next section. Using various text mining techniques based on the R data language and its text mining techniques [20–22], we preprocess the collected patent documents and construct the patent–keyword matrix in Steps 2 and 3. The 'tm' package in our research is one of the R text mining packages and the build the document-term matrices from the documents using data import, a document corpus, and natural language processing, such as stemming, whitespace elimination, stop-word removal, etc. [21]. In general, the number of keywords in the patent–keyword matrix is too large for Bayesian network modeling. To solve this problem, we performed the factor analysis and reduced the dimension of the patent–keyword matrix. In this paper, we use R data language and its factor analysis function [19]. Next, we carry out the Bayesian network modeling according to the selected factors. In this analysis, we considered the 'bnlearn' package based on R data language [4,7,20]. Finally, we found the technological structure to understand the DAI technology and forecast the future state of DAI technology using the results of factor analysis and Bayesian network modeling. The results of the DAI technological structure are used to build the R&D strategy for DAI technology management.

4. Case Study Using DAI Patent Data

To illustrate how our proposed method can be applied to real domains, we carried out a case study using the patent documents related to the DAI extracted from the patent databases in the world [18,19]. We used the following keyword searching equation to retrieve the patent documents for DAI:

> *(((Artificial and intelligen *) or (deep and learn *) or (machine and learn *) or (data and analysis) or statistic *) and (disaster or environment or geoscience or "remote sensor" or Atmosphere or Cryosphere or Ocean or Land or hazard or climate or earthquake or fire or burn or blackout or blow or "wind cloud" or arson or arsonist or avalanche or barometer or "Beaufort scale" or blackout or blizzard or blow or cloud or crust or "cumulonimbus" or cyclone or dam or drought or "dust storm" or earthquake or erosion or fatal or fault or fire or flood or fog or force or forest or "forest fire" or gale or geyser or gust or hail or hailstorm or heat or high-pressure or hurricane or iceberg or kamikaze or lack or lava or lightning or "low pressure" or magma or mountain or nimbus or ocean or permafrost or rain or rainstorm or Richter scale or river or sandstorm or sea or seismic or sinking or snowstorm or storm or stuck or thunderstorm or tornado or tsunami or twister or "violent storm" or volcano or volt or whirlpool or whirlwind or "wind scale" or "wind vane" or "windstorm" or "heat wave" or wave or tremor or underground or death or casualty or fatality or disaster or money or lost or damage or life or poor or shelter or rescue or coast or monster or myth or science or scientist or god or goddess or sink or boat or destruction or destroy or uproot or tree or fate or poverty or impoverish or farm or touchdown or zap or tension or nightmare or monstrosity or oil or spill or cataclysm or Bermuda or "Bermuda Triangle" or wind or windy or wave or ice))*

We had collected the DAI patent documents filled with patent offices around the world by 2018. Finally, we selected 16,875 valid patents and used them in the case study of this paper. In addition,

we extracted 251,358 terms from the patent documents using R data language and its packages [20–22]. So, we made a matrix with 16,875 rows (patents) and 251,358 columns (terms) as a structured set of data for Bayesian network modeling and factor analysis. From the 251,358 terms, we selected the following 162 patent keywords representing DAI with the help of domain experts of DAI [23]: "abnormal", "acoustic", "air", "alarm", "amplitude", "analysis", "antenna", "artificial", "audio", "automatic", "band", "battery", "beam", "big", "cable", "camera", "car", "channel", "cloud", "cluster", "coal", "communication", "computing", "cylinder", "damage", "data", "database", "deep", "depth", "detection", "device", "diagnosis", "digital", "display", "earth", "earthquake", "echo", "edge", "electric", "energy", "engine", "engineering", "environment", "estimation", "fault", "feedback", "fire", "flow", "fluid", "forecast", "frame", "fuzzy", "gas", "geological", "grid", "health", "hole", "human", "image", "information", "intelligence", "interaction", "interface", "land", "language", "laser", "layer", "learning", "life", "light", "lightning", "liquid", "machine", "magnetic", "map", "measurement", "memory", "metal", "mobile", "monitoring", "natural", "negative", "network", "neural", "node", "normal", "oil", "optical", "parallel", "patient", "pattern", "physical", "picture", "pipe", "pipeline", "pixel", "plane", "platform", "power", "prediction", "pressure", "probability", "protection", "protocol", "pulse", "pump", "radar", "radio", "remote", "risk", "road", "robot", "rock", "sampling", "satellite", "scale", "scanning", "sea", "security", "seismic", "sensor", "signal", "software", "soil", "space", "spatial", "speed", "stability", "statistics", "steel", "stream", "surface", "switch", "tank", "temperature", "time", "transmission", "tree", "tunnel", "turbine", "ultrasonic", "underground", "user", "valve", "vehicle", "velocity", "video", "virtual", "visual", "voice", "voltage", "warning", "water", "wave", "waveform", "wavelet", "weather", "web", "wheel", "wind", "wire", and "wireless". Using these keywords, we constructed the analytical matrix as shown in Figure 2.

	abnormal	acoustic	air	⋯	wind	wire	wireless
Patent 1							
Patent 2							
⋮							
Patent 16875							

V_{ij} : *occurred frequency value of jth keyword in ith patent*

Figure 2. Analytical matrix for Bayesian network modeling and factor analysis.

The analytical matrix consists of 162 keywords and 16,875 patents, and each element of the matrix has an occurred frequency value of a keyword in each patent. First, we performed factor analysis using this matrix. In this paper, we selected the top 10 factors with high factor loadings from the results of the factor analysis. Table 1 shows the representative keywords included in the top 10 factors.

Table 1. Top 10 factors and representative keywords.

Factor	Keywords
1	Air, coal, cylinder, device, flow, fluid, gas, hole, liquid, monitoring, oil, pipe, pipeline, pressure, pump, rock, sensor, tank, temperature, underground, valve, water
2	Energy, power, prediction, speed, turbine, wind
3	Alarm, analysis, big, cloud, communication, computing, data, database, device, display, environment, health, information, interaction, interface, machine, mobile, monitoring, network, platform, protocol, remote, security, sensor, stream, transmission, user, video, virtual, warning, wave, wireless
4	Abnormal, alarm, analysis, battery, communication, data, diagnosis, electric, energy, fault, forecast, grid, information, lightning, monitoring, network, power, prediction, protection, signal, stability, switch, time, transmission, voltage, warning, weather
5	Band, camera, detection, digital, edge, frame, image, picture, pixel, remote, scale, video, visual
6	Amplitude, analysis, big, data, depth, earth, earthquake, geological, layer, road, rock, seismic, speed, surface, time, velocity, wave, waveform, wavelet
7	Artificial, intelligence, neural, robot

Table 1. *Cont.*

Factor	Keywords
8	Audio, channel, communication, detection, device, digital, forecast, grid, interface, map, prediction, pulse, radio, sensor, signal, space, spatial, switch, transmission, ultrasonic, voltage, wave, waveform, weather, wireless
9	Deep, depth, layer, learning, machine, network, neural, node, prediction
10	Beam, camera, detection, echo, laser, light, measurement, optical, plane, radar, scanning, surface, tunnel

Using the results in Table 1, we performed a Bayesian network analysis to identify the technological association between the keywords representing each factor. Figure 3 shows the relationship between the technological keywords in factors 1 and 2.

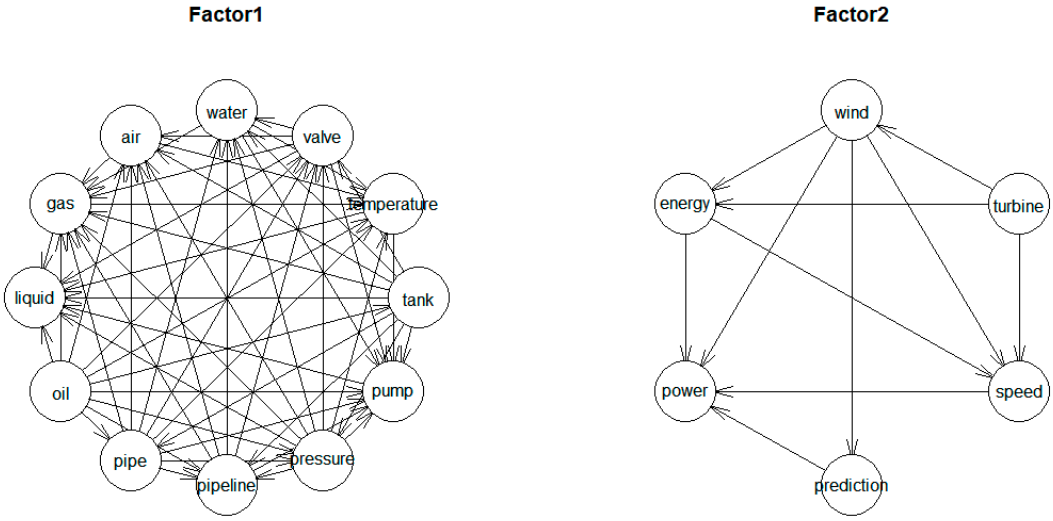

Figure 3. Bayesian networks of the technological keywords in factors 1 and 2.

Factor 1 is the factor that contains the 12 most technological keywords. Using the keywords in factor 1, we can define factor 1 as representing the technology of the 'control of air and water'. Also, in the factor 1, the technologies related to value and pump are important because the nodes (value and pump) receive numerous arrows from various nodes. The same applied to the temperature node. In addition, factor 2 in Figure 3 represents the technology of 'natural energy' according to the keywords included in factor 2. The meaningful keywords in factor 2 are wind and speed because they are connected to many other nodes. We show the Bayesian network results of factors 3 and 4 in Figure 4.

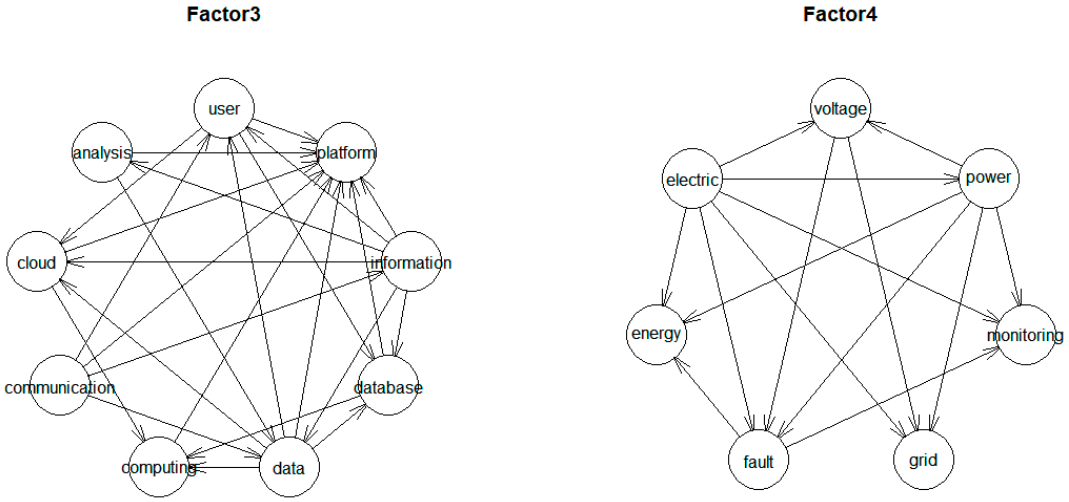

Figure 4. Bayesian networks of the technological keywords in factors 3 and 4.

We defined the representative technology of factor 3 as a 'database and data analysis platform' using the 9 keywords in factor 3. The keywords data and platform are important keywords in factor 3. So, we found that the big data platform for DAI is necessary to manage the various disasters efficiently. We also determined that the technology represented by factor 4 is 'electric energy' by the 7 keywords in factor 4. Next, the results of Bayesian network modeling for factors 5 and 6 are shown in Figure 5.

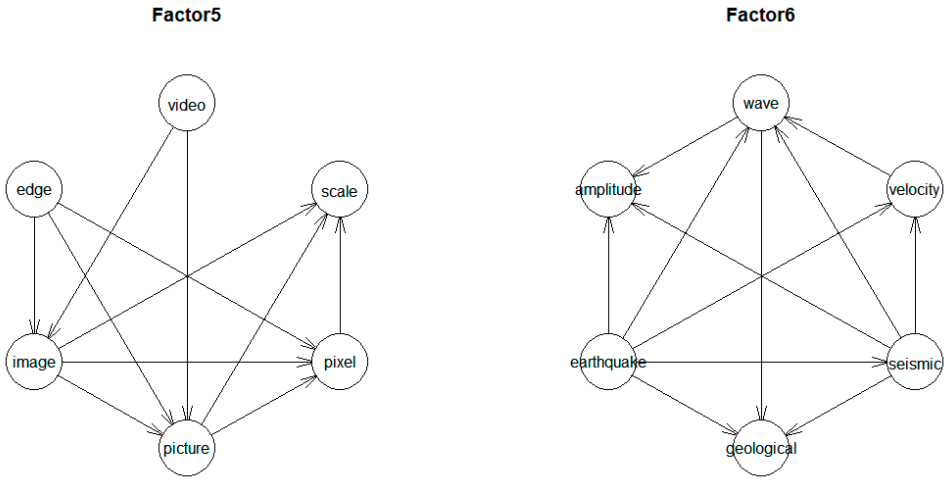

Figure 5. Bayesian networks of the technological keywords in factors 5 and 6.

Using the 6 technological keywords in factor 5, we can define the representative technology of factor 5 as 'image data management'. In other words, this shows the importance of image data in DAI. The technology represented by factor 6 is related to 'earthquake'. The fact that earthquakes represent one factor in DAI suggests that earthquakes are a big part of entire disasters. In other words, effective management of earthquakes is a very important part of DAI. In addition, the keywords of wave and geological have many connections to other keywords. This means that we should consider the technologies related to wave and geological with earthquake technology. Figure 6 shows the Bayesian network modeling results of factors 7 and 8.

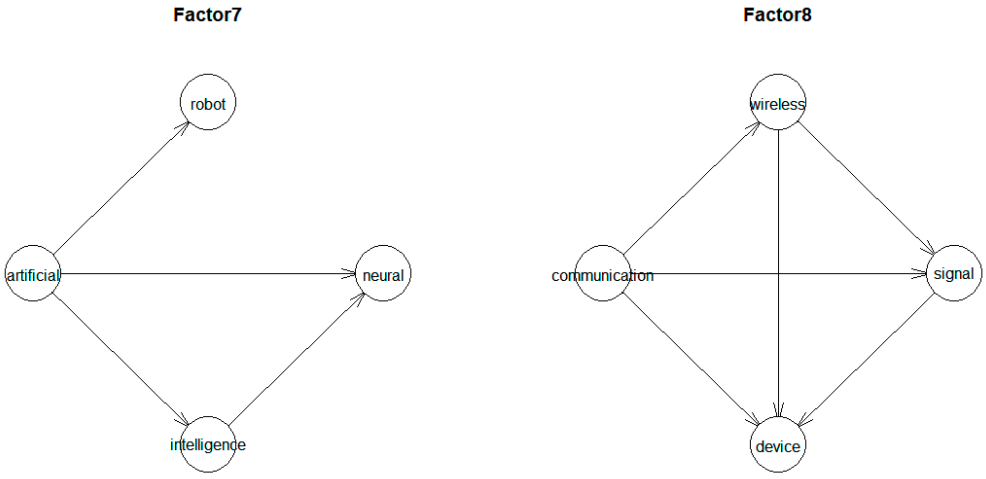

Figure 6. Bayesian networks of the technological keywords in factors 7 and 8.

Factor 7 contains the technology of 'AI and robot', and neural networks are also an important issue explained by factor 7. The representative technology of factor 8 is the technology related to 'signal and communication device'. The wireless technology is also important in factor 8 technology. We illustrate the results of the Bayesian networks related to factors 9 and 10 in Figure 7.

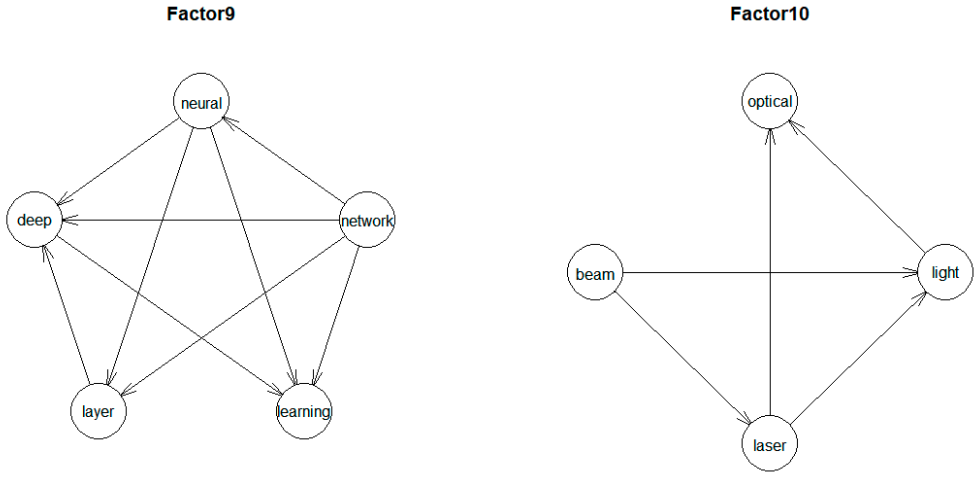

Figure 7. Bayesian networks of the technological keywords in factors 9 and 10.

We defined the technology representing factor 9 as 'machine learning and deep learning' technology. In addition, the factor 9 contains the technology of neural networks, and this is similar to factor 7 in Figure 6. Lastly, factor 10 is defined as the technology of 'optical and light'. This technology is different to the previous factors, but using the results of the DAI patent analysis, we found that the technology related to 'optical and light' is necessary for DAI. Putting together the analysis results of the Bayesian models of Figures 3–7, we defined the technologies represented by the top ten factors that explain DAI in Table 2.

Table 2. Top 10 factors and representative technologies.

Factor	Representative Technology
1	Control of air and water
2	Natural energy
3	Database and data analysis platform
4	Electric energy
5	Image data management
6	Earthquake
7	AI and robot
8	Signal and communication device
9	Machine learning and deep learning
10	Optical and light

Five of the 10 factors (factors 3, 5, 7, 8, 9) were related to data analysis, machine learning, and intelligent systems. In other words, half of all technologies for DAI are related to learning from data. The number of factors related to energy is two (factors 2, 4). Factor 1 explains the technology of the control of air and water, and factor 10 shows the optical and light technology. Unlike other factors, factor 6 represents an earthquake. One factor explains only one earthquake. This means that earthquakes are as influential in DAI as other factors. Finally, Bayesian network modeling, using factors as nodes, was performed to identify the relationship between the technologies represented by each factor in DAI. Figure 8 illustrates the result of Bayesian networks modeling using the 10 factors.

The factors related to energy (factors 2 and 4) are connected to various other factors. This means that the energy technology is important to DAI. We have identified that the core technology of DAI is factor 9 (machine learning and deep learning). It can also be seen that Factors 3 (database and data analysis platform) and 7 (AI and robot) support this technology. Factor 9 also affects factor 8 (signal and communication device) and factor 10 (optical and light), contributing to the efficient operation of DAI. Factor 10 influences factor 5 (image data management) in DAI. The technologies affecting factor 6

(earthquake) were identified as factors 2 (natural energy), 3, and 10. We leave the more meaningful interpretations and applications of Figure 8 to an expert in DAI.

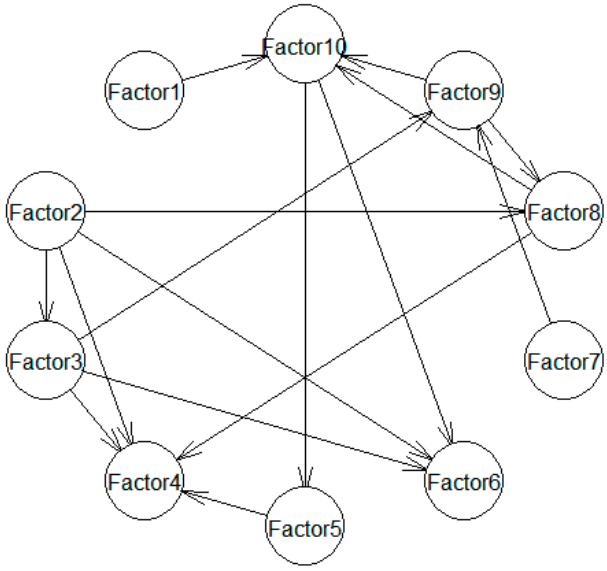

Figure 8. Bayesian network of the top 10 factors.

5. Conclusions

In this paper, we proposed a method to analyze patent data for technology analysis of the target domain. We combined Bayesian network modeling with factor analysis to make our proposed model. In addition, we applied the patent documents related to DAI to the proposed model. To perform the proposed model based on machine learning and statistics, we constructed a structured set of data from the collected patent documents. We extracted the patent keywords from the patent data and made a patent–keyword matrix with an occurred frequency of keywords in the patent. After the preprocessing by text mining, we got the patent–keyword matrix with 16,875 patents and 162 keywords. First, using factor analysis, we selected 10 factors from the 162 keywords. In our research, each factor represents its corresponding technology. For example, factor 9 of this paper illustrates the technology of machine learning and deep learning. Finally, we performed Bayesian network modeling using 10 factors to identify the relationships among sub-technologies for DAI.

In the previous studies, the factor analysis and Bayesian network modeling were actively used in diverse areas, respectively. In this paper, we proposed a new patent analysis methodology by combining two different methods (factor analysis and Bayesian network modeling). In particular, the most important contribution of this study is providing new technology analysis results in the DAI field using the statistics and machine learning algorithms.

In our future works, we will consider more advanced approaches related to Bayesian inference and learning for patent data analysis. Furthermore, we will use more diverse information extracted from patent documents such as citations, IPC codes, claims, applied and registered dates, etc. Specifically, we are to develop a learning model using prior likelihood and a posterior of Bayesian statistics. In addition, we will consider hierarchical, nonlinear, and nonparametric Bayesian models to find the meaningful relations between technologies. Therefore, we will provide various methodologies for patent technology analysis based on Bayesian inference and learning. In addition, we will not only analyze one technological field (the DAI in this paper), but we will also study the methodology that can perform an analysis of technological association in several technological fields, such as bio and AI at the same time. We will consider the dynamic Bayesian networks for this [24].

Author Contributions: S.J. designed this study and collected the data for the experiment. S.P. preprocessed the data and selected valid patents and analyzed the data to show the validity of the study and wrote the paper and performed all the research steps. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

1. Roper, A.T.; Cunningham, S.W.; Porter, A.L.; Mason, T.W.; Rossini, F.A.; Banks, J. *Forecasting and Management of Technology*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
2. Choi, J.; Jun, S.; Park, S. A patent analysis for sustainable technology management. *Sustainability* **2016**, *8*, 688. [CrossRef]
3. Park, S.; Jun, S. Statistical Technology Analysis for Competitive Sustainability of Three Dimensional Printing. *Sustainability* **2017**, *9*, 1142. [CrossRef]
4. Scutari, M. Learning Bayesian Networks with the bnlearn R Package. *J. Stat. Softw.* **2010**, *35*, 1–22. [CrossRef]
5. Korb, K.B.; Nicholson, A.E. *Bayesian Artificial Intelligence*, 2nd ed.; CRC Press: London, UK, 2011.
6. Nagarajan, R.; Scutari, M.; Lebre, S. *Bayesian Networks in R with Application and System Biology*; Springer: London, UK, 2013.
7. Scutari, M. Bayesian Network Constraint-Based Structure Learning Algorithms: Parallel and Optimised Implementations in the bnlearn R Package. *J. Stat. Softw.* **2017**, *77*, 1–20. [CrossRef]
8. Theodoridis, S. *Machine Learning a Bayesian and Optimization Perspective*; Elsevier: London, UK, 2015.
9. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2012.
10. Hunt, D.; Nguyen, L.; Rodgers, M. *Patent Searching Tools & Techniques*; Wiley: Hoboken, NJ, USA, 2007.
11. Kim, J.; Sun, B.; Jun, S. Sustainable Technology Analysis Using Data Envelopment Analysis and State Space Models. *Sustainability* **2019**, *11*, 3597. [CrossRef]
12. Kim, J.; Yoon, J.; Hwang, S.Y.; Jun, S. Patent Keyword Analysis Using Time Series and Copula Models. *Appl. Sci.* **2019**, *9*, 4071. [CrossRef]
13. Kim, J.; Jun, S.; Jang, D.; Park, S. An Integrated Social Network Mining for Product-based Technology Analysis of Apple. *Ind. Manag. Data Syst.* **2017**, *117*, 2417–2430. [CrossRef]
14. Jun, S. IPC code Analysis of Patent Documents Using Association Rules and Maps—Patent Analysis of Database Technology. *Commun. Comput. Inf. Sci.* **2011**, *258*, 21–30.
15. Kim, K.H.; Han, Y.J.; Lee, S.; Cho, S.W.; Lee, C. Text Mining for Patent Analysis to Forecast Emerging Technologies in Wireless Power Transfer. *Sustainability* **2019**, *11*, 6240. [CrossRef]
16. Dumais, S.T. Latent semantic analysis. *Annu. Rev. Inf. Sci. Technol.* **2004**, *38*, 188–230. [CrossRef]
17. Johnson, R.A.; Wichern, D.W. *Applied Multivariate Statistical Analysis*, 6th ed.; Pearson: Essex, UK, 2012.
18. WIPSON. WIPS Corporation. Available online: http://www.wipson.com (accessed on 15 December 2018).
19. USPTO. The United States Patent and Trademark Office. Available online: http://www.uspto.gov (accessed on 15 December 2018).
20. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019; Available online: http://www.R-project.org (accessed on 9 September 2019).
21. Feinerer, I.; Hornik, K. Package 'tm' Ver. 0.7–5, Text Mining Package, CRAN of R Project. 2018. Available online: https://cran.r-project.org/web/packages/tm/tm.pdf (accessed on 1 January 2019).
22. Feinerer, I.; Hornik, K.; Meyer, D. Text mining infrastructure in R. *J. Stat. Softw.* **2008**, *25*, 1–54. [CrossRef]
23. KISTA, Korea Intellectual Property Strategy Agency. 2019. Available online: https://cran.r-project.org/web/packages/arm/arm.pdf (accessed on 1 March 2019).
24. Fröhlich, H.; Bahamondez, G.; Götschel, F.; Korf, U. Dynamic Bayesian Network Modeling of the Interplay between EGFR and Hedgehog Signaling. *PLoS ONE* **2008**, *10*, e0142646. [CrossRef] [PubMed]