Article

Patent Analysis Using Bayesian Data Analysis and Network Modeling

Sangsung Park and Sunghae Jun *

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

Abstract: Patent analysis is to analyze patent data to understand target technology. Patent data contains various detailed information about the developed technology. Therefore, many studies concerning patent analysis have been carried out in the technology analysis fields. Most traditional methods for technology analysis were based on qualitative approaches such as Delphi survey. However, the patent analysis methods based on statistics and machine learning have been introduced recently. In this paper, we proposed a statistical method for quantitative patent analysis. Moreover, we selected drone technology as the target technology for patent analysis. To understand drone technology, we analyzed the patents on drone technology. We searched the patent documents related to drone technology and transformed them to structured data using text mining techniques. First, we visualized the patent keywords to identify the technological structure of a drone. Next, using Bayesian additive regression trees, we analyzed the structured patent data to construct technology scenarios for drones. To illustrate the performance and validity of our proposed research, we presented the experimental results of patent analysis using patent documents related to drone technology.

Keywords: Bayesian additive regression trees; keyword visualization; management of technology; multiple linear regression; technology analysis

1. Introduction

Patent analysis is to analyze the patent documents related to target technology using data analysis methods based on statistics and machine learning. Moreover, the aim of patent analysis is to understand the target technology using the results of patent document analysis. In this paper, we selected drone technology, which is rapidly spreading recently, as a target technology for the proposed patent analysis method. A drone is an unmanned aerial vehicle that flies automatically or semi-automatically by a program without a pilot on board [1]. Many companies in the world are conducting various R&D activities to expand and commercialize drone system in diverse fields such as logistics, disaster relief, national defenses, and entertainment [2–7]. Tejado-Ramos et al. (2021) used drone remote sensing technology to enhance the sustainability of wolfram mining. They performed an effective exploration of the mine using the drone remote sensing technology [6]. In addition, Suzuki (2018) studied on the use of drone technologies in robotics area. The author carried out the research on structure and mechanism of drone system [5]. As the demand of the public sector as well as the private continues to expand with the development of unmanned aerial vehicle technology, the commercial use of drones and the demand for personal flying vehicles are expected to increase [1,8]. As the drone technology is increasingly coming into our daily lives, we need a deeper understanding of this technology. Nouacer et al. (2020) emphasized technological importance of object detection and avoidance, management of air traffic, control, and security for the development of drone

system. They also illustrated the convergence of drone technologies and markets in Europe [1]. In order to improve our understanding of drones, we have to conduct a drone technology analysis. For drone technology analysis, we use the patents related to drone technology searched from the world patent databases, because the patents contain most information about the developed technology [9,10]. However, the studies that quantitatively analyze drone technology using statistical patent analysis have not yet been actively conducted. Thus, in this paper, we carried out drone patent analysis using statistical modeling for technology analysis of drone systems. For the patent analysis by statistics and machine learning algorithms, it is necessary to preprocess the patent documents for constructing structured data such as the table of relational databases [11,12]. The table structure consists of row and column for observation (document) and variable (keyword), respectively. Thus, we built patent-keyword matrix as a structured data by text mining techniques. Using the structured patent data, we visualized the patent keywords by word cloud and correlation networks, and analyzed them by Bayesian additive regression trees (BART) for the drone patent analysis. From the visualization and analysis results of patent data related to drone technology, we identified the technological relationship of drone technology and used it for drone technology management. To this purpose, we finally created technology scenarios for drone technology development such as technology scenarios. The technology scenarios contribute to technological deployment of drones such as R&D planning, new product and service development, technological innovation, and technology forecasting of drone systems. To verify the performance and validity of our research, we showed how the proposed methodology can be applied to the practical domain using the patent documents collected from the patent databases such as the United States Patent and Trademark Office (USPTO), WIPS Corporation (WIPSON), and the Korea Intellectual Property Rights Information Service (KIPRIS) [13–15]. Our research contributes to the objective approach towards patent analysis using quantitative methods for technology analysis. Compared to subjective technology analysis that relies on the knowledge and judgment of expert groups such as Delphi survey [10–12], our proposed method focuses on quantitative patent analysis for objective technology analysis using patent data. As big data processing and analysis methods develop, the use of quantitative technology analysis using patent big data increases. In such a big data environment, we expect that the proposed method will contribute to the field of technology management as one of several objective technology analysis methods.

We organize this paper as follows. In Section 2, we introduce the existing studies related to our study such as drone technology and patent big data analysis. We propose a method of drone technology analysis using Bayesian additive regression trees and correlation networks in Section 3. In the next section, we illustrate performance of our proposed method for drone technology analysis. In the last section, we discuss the conclusion of this paper and future research tasks.

2. Patent Data Analysis

Big data are characterized by the volume and heterogeneity of data [16]. With the development of hardware as well as software, the size of data that can be stored in computer repositories is continuously increasing. The data included in big data consist of very diverse types such as numbers, texts, images, video, and symbols. The patent document data also have the same characteristics of big data [17,18] because a patent document consists of various data types such as the title, inventor, application date, specification, figures, drawings, and citation information [9,10,19]. In addition, the number of patents applied and registered with the patent offices around the world is very large. Therefore, we can analyze patent document data using the existing big data analysis methods provided by statistics and machine learning algorithms. The analysis methods require a structured data type as input data before patent analysis. The data have a structure similar to table in relational database [20,21]. The table is a matrix with patent documents and keywords for rows and columns, respectively. The element of this matrix is frequency of an occurred

keyword in a patent document. We carried out technology analysis using this patent–keyword matrix. In this paper, we built this matrix using the collected patent documents related drone technology and analyzed the matrix data for drone technology analysis. We show a comparison of existing studies related to patent data analysis in Table 1.

Table 1. Comparison of patent data analysis.

Reference	Technology Field	Methodology	Analysis Purpose
11	Liquid crystal display wide viewing angle	Text mining, factor analysis, morphological analysis	Evaluating morphological configuration
12	Artificial intelligence	Text mining, statistical inference, Bayesian statistics	Technology forecasting
17	Smart devices of Apple	Superpopulation model, linear regression	Technology structure
18	Three dimensional printing	Text mining, fuzzy regression, fuzzy clustering	Technology clustering, R&D planning
19	General technology	Citation network analysis, exponential random graph models	Technology citation and network

In Table 1, we illustrate reference, technology field, methodology, and analysis purpose of previous research work for patent analysis. We found that the existing patent analyses were performed in various technology fields. Moreover, we knew that the purpose of technology analysis is widely distributed from R&D planning to technology forecasting. On the other hand, most technology analysis methodologies were based on text mining, statistics, and machine learning.

Before conducting patent analysis on drone technology using the proposed method, we surveyed the drone technology on four perspectives: strength, weakness, opportunity, and threat (SWOT) from the previous research [1–8]. On the basis of the SWOT analysis result, we extracted the drone technology scenarios from the result of drone patent analysis. Table 2 shows the result of SWOT analysis of drone technology.

Table 2. SWOT of drone technology.

SWOT	Description
Strength	Aircraft manufacturing and communication technologies, and their infrastructure Willingness to develop drones at national level
Weakness	Lack of intelligent software technology for drone operation Exclusive drone technology focused on developed countries
Opportunity	Continuous growth of drone market Expansion of drone services linked to other industries
Threat	Possibility of using drones that endanger human privacy Use of drones as weapons

In the results of SWOT analysis, the strengths of drone technology are the technologies and infrastructures of aircraft manufacturing and communication and the willingness to develop drones at national level. On the other hand, the weaknesses of drone technology are the lack of intelligent software technology for drone operation and the exclusive

drone technology focused on developed countries. Moreover, the opportunity of drone technology is based on continuous growth of drone market and expansion of drone services linked to other industries. Finally, the threat of drone technology is dependent on possibility of using drones that endanger human privacy and use of drones as weapons. Therefore, on the basis of the SWOT result shown in Table 2, we analyzed the patent data of drones. The technologies for building drone systems and services have been developed rapidly. In addition, the size of the drone market is growing at a very fast pace. However, in reality, drones have various factors that threaten humans. Above all, it is important to know the drone technology with the potential to threaten human privacy and safety. On the basis of the result of SWOT on drone technology, we carried out the patent analysis of drone technology.

3. Proposed Method

3.1. Technical Analysis Procedure

For drone technology analysis, we analyzed the patent document data related to drone technology using patent keyword visualization and BART modeling. The following procedure represents all steps of our drone technology analysis from collecting patent documents to performing patent data analysis.

(Step 1)   Collecting Drone Patent Documents

(1-1) Searching patent documents related to drone technology from patent databases.
(1-2) Removing noise from searched patent documents to select valid patents.

(Step 2)   Preprocessing Valid Patent Data

(2-1) Representing collection of text documents from valid patent data.
(2-2) Parsing text collection to build text database (DB).
(2-3) Constructing the patent–keyword matrix as structured data.

(Step 3)   Visualizing Patent Keywords

(3-1) Generating word cloud for finding top frequency keywords.
(3-2) Making correlation networks between all keywords.
(3-3) Calculating degree values of all keywords for ranking of core keywords.

(Step 4)   Performing Bayesian Additive Regression Trees (BART) model

(4-1) Comparing *p*-values of explanatory keywords using a multiple regression model.
(4-2) Selecting important keywords using BART modeling.
(4-3) Building technology scenarios of drone using all results.

Our proposed method consists of four steps from searching drone patents to building technology scenarios. In Step 2, we constructed a patent–keyword matrix using text-mining techniques. The visualization of word cloud and correlation networks requires this matrix data. In addition, the multiple regression and BART models need the matrix type based on table structure of DB. Using the results of Steps 3 and 4, we selected the important keywords for developing drone technology. In step (4–3), we built technology scenarios of drones using the results of visualization and analytical modeling. The technology scenarios can be applied to technology management of drones such as R&D planning and new service development of the drone system.

3.2. Text Mining

We used various text mining techniques to transform patent documents into structured data [21–23]. In this paper, we also used the text mining methods to build the structured data from the collected patent documents. We show the general process of patent data analysis in Figure 1.
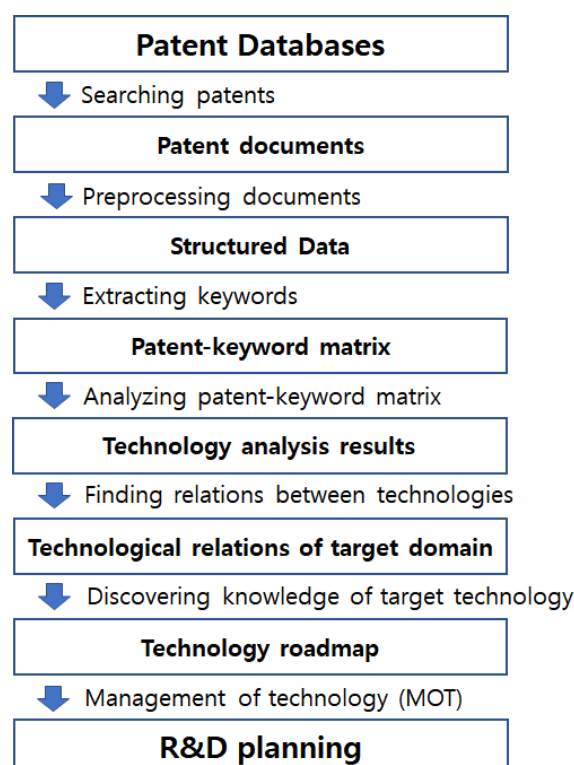
Figure 1. Process of patent data analysis.

The start of patent data analysis is to collect the patent documents related to target technology from the world patent databases. Next, we preprocessed the patent documents to build structured data and extracted technological keywords from the patent data to construct the patent–keyword matrix. Using statistics and machine learning algorithms, we analyzed the matrix data to obtain the results of technology analysis. The results are used for us to find the technological relations between core technologies in target domain. We discovered the knowledge about target technology and built technology scenarios for management of technology (MOT) such as R&D planning. In our research, we also followed the general process of patent data analysis. In addition, our target domain is the drone technology, and we propose patent analysis and visualization methods using the BART and correlation networks.

We used the patent documents related to drone technology to find technology structure between core technologies of drones. The proposed technology analysis is based on statistics and machine learning algorithms. Most methods of statistics and machine learning require a structured data type for data analysis. A patent is not structured data, it is a document containing text, dates, numbers, pictures, etc. Thus, we had to transform the patent documents into the matrix data. Among the variety of data included in patent documents, we used text data with developed technology information. The rows and columns of this matrix represent patent documents and keywords, and the matrix elements are frequency values of keywords occurred in patent document.

3.3. Structured Data

To build the structured data that is the document–keyword matrix of Section 3.1, we preprocessed the patent documents using text mining techniques. Text mining is defined as a text preprocessing and computing for text analysis [21–23]. Figure 2 shows the process of making structured data using text mining from the patent documents.
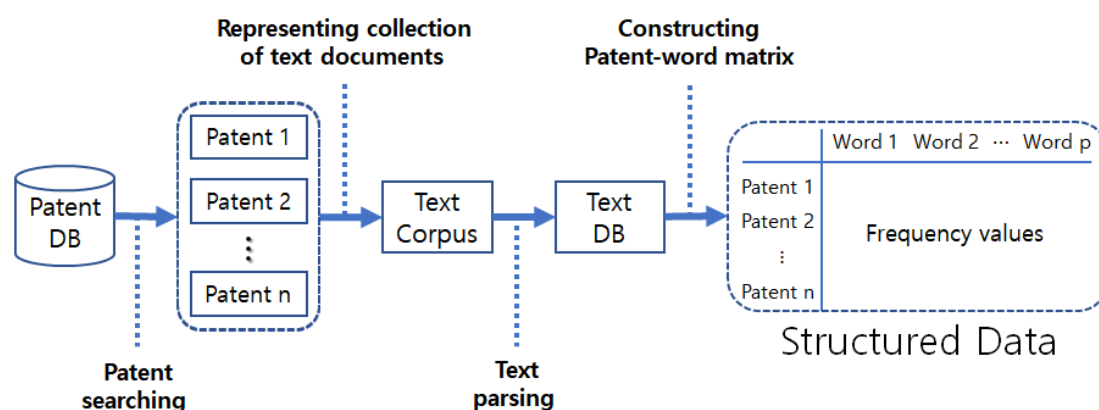
Figure 2. Constructing structured data using text mining.

We searched the patent documents from the patent repositories such as the USPTO, KIPRIS, and WIPSON [13–15]. We made a keyword equation of drone technology for the patent searching. Through noise filtering on the searched patent documents, we finally obtained n valid patent documents. The documents were changed to text corpus by representing collection of patent documents. The text corpus consisted of various words representing all patent documents. Using grammatical parsing and preprocessing, we transformed the text corpus to semi-structured form (text database). Finally, we created the patent–keyword matrix as structured data from patent text database. This structured data are used in various analyses provided by statistics and machine learning algorithms. Figure 3 shows a part of the matrix data.

```
     aerial control connect devic system flight aircraft wing arrang power data rotor rotat motor drive
1       0      0      4      0     1      0      1      1     0     0    0    0     0     0     0
2       0      0      4      0     1      0      1      1     0     0    0    0     0     0     0
3       0      0      0      2     3      1      6      3     2     1    0    0     2     0     0
4       8      0      3      3     0      0      0      0     1     0    0    0     0     0     0
5       8      0      3      3     0      0      0      0     1     0    0    0     0     0     0
6       0      2      0      1     0      0      0      0     0     0    0    0     1     0     0
     camera signal detect air time batteri sensor ground direct bottom shaft propel fuselag monitor drone
1       0      0      0     0    0      0      0      0      0      0     0     0      4       0     0
2       0      0      0     0    0      0      0      0      0      0     0     0      4       0     0
3       0      0      0     2    0      0      0      0      0      0     0     1      3       0     0
4       2      3      0     0    0      0      0      0      0      1     0     0      0       0     0
5       2      3      0     0    0      0      0      0      0      1     0     0      0       0     0
6       0      0      0     0    0      0      0      0      0      0     0     0      0       0     0
     plane automat wireless charg area station remot measur gear speed machin water storag image light
1      0      0       0      0    0     0      0     0      0     0     0     0     0      0     0
2      0      0       0      0    0     0      0     0      0     0     0     0     0      0     0
3      0      0       0      0    0     0      0     0      0     2     0     0     0      0     0
4      0      2       0      0    0     0      0     0      0     0     0     0     0      0     0
5      0      2       0      0    0     0      0     0      0     0     0     0     0      0     0
6      0      0       1      0    0     0      1     0      0     0     0     0     0      0     5
     circuit tail space video mobil network stabil gps interfac weight electron pressur wind autonom map
1      0      0     0    0     0      0       0    0      0       1       0       0     0      0     0
2      0      0     0    0     0      0       0    0      0       1       0       0     0      0     0
3      0      0     0    0     0      0       0    0      0       0       0       1     0      0     0
4      0      0     0    0     0      0       0    0      0       0       0       0     0      0     0
5      0      0     0    0     0      0       0    0      0       0       0       0     0      0     0
6      1      0     0    0     1      0       0    0      0       0       0       0     0      0     0
```

Figure 3. Structured data of drone patent documents.

We illustrate only a portion (six patent documents) of all patents. Moreover, it is the result of selecting major 60 keywords among the words appearing in the entire drone patent document.

3.4. Keyword Visualization

In this paper, we used the R data language and packages for preprocessing patent data by text mining [23,24]. R provides programming environment for data analysis and visualization [24]. Moreover, the package *tm* is a popular R package for text mining [22,23]. We used this package to create structured data through preprocessing of the drone patent documents in Figure 2. Next, we analyzed the structured data using visualization and BART modeling. The proposed method for drone technology analysis is composed of two approaches. The first approach is the keyword visualization with word cloud and correlation networks. The second is statistical modeling based on multiple linear regression and BART models. We first carried out word cloud to show the relative strength of patent keywords representing drone technology. A word cloud is a method to visualize the intensity of keyword as the size of text on the cloud shape [25,26]. Using the word cloud result, we selected the keywords that had a strong influence on drone technology development. Next, we performed correlation analysis on keywords. The correlation coefficient between keywords X and Y is defined as (1) [20].

$$\text{Cor}(X, Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)S_X S_Y} \tag{1}$$

where $\bar{X}$ and $\bar{Y}$ are mean values of occurred frequency values of X and Y, respectively. n is the number of collected patent documents. Moreover, S_X and S_Y are standard deviations of X and Y, respectively. We obtained correlation coefficient values for all keyword pairs and used them for visualization of correlation networks. This was based on graph data structure consisting of node and edge. The node and edge were keyword and connection between two keywords, respectively. In the proposed correlation networks, the edge is connected when the correlation coefficient value between two keywords is greater than or equal to the threshold. Figure 4 illustrates the structure of correlation networks.
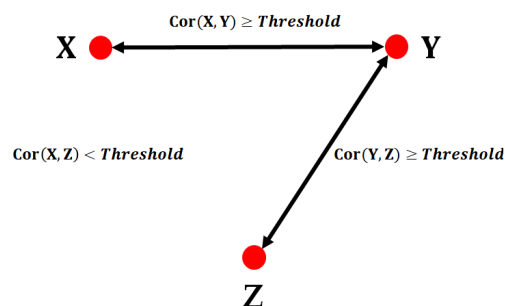

Figure 4. Correlation network structure.

In Figure 4, the network contains three nodes (keywords)—X, Y, and Z. The edge of X and Y is connected because the correlation coefficient value between X and Y is greater than or equal to the threshold value. For the same reason, the keywords Y and Z are also connected. However, the edge of keywords X and Z is not connected because their correlation coefficient value is less than the threshold. To find the technological structure for drone technology, we perform the visualization of correlation networks using all patent keywords. The threshold value is predetermined before visualization. In our experiments, we compared the visualization results according to the change of threshold values. In addition, we considered only undirected graph in our networks. The undirected graph has a bidirectional network structure that connects two keywords to each other. Thus, the degree of the keyword with undirected graph is counted as twice the number of other linked keywords because the degree is the sum of indegree and outdegree. The degree of node v is defined as Equation (2).

$$\text{degree}(v, G) = N^+(v) + N^-(v) \tag{2}$$

where G is the network including node v. $N^+(v)$ and $N^-(v)$ are indegree and outdegree of v, that is, the indegree represents the number of connecting lines coming from other nodes in the same network to node v. Conversely, the outdegree illustrates the number of connecting lines from node v to the other nodes. Therefore, the degree of node v is calculated as the sum of indegree and outdegree. For example, the degree of keyword X is 2 because X is connected to one keyword (Y) in Figure 4. In addition, the degree of keyword Y is 4 because Y is linked to two keywords (X and Z). The degree results of keywords are used to select the meaningful keywords for drone technology.

3.5. BART Modeling

Next, we analyze the drone patent data using statistical models. Thus, we use multiple linear regression and BART for the patent analysis of drone technology. In the multiple regression model of Equation (3), the response variable Y is the keyword *drone* and the other keywords except *drone* are used as explanatory variables $(X_1, X_2, \ldots, X_p)$.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon, \ i = 1,2,\ldots,n \tag{3}$$

where Y and $(X_1, X_2, \ldots, X_p)$ represent the frequency values of patent keywords occurring in each patent document. ε is a noise term distributed to normal distribution with mean = 0 and variance = σ^2. Moreover, β_0 is an intercept of regression equation, and $(\beta_1, \beta_2, \ldots, \beta_p)$ are the parameters corresponding to input keywords (explanatory variables). To find X_i that has a statistically significant effect on Y, we perform the following hypothesis test.

$$H_0: \beta_i = 0 \quad vs. \quad H_1: \beta_i \neq 0 \tag{4}$$

H_0 represents the null hypothesis in that the regression coefficient β_i of X_i becomes 0. This means that the explanatory variable X_i has no significant effect on Y. On the other hand, H_1 represents the alternative hypothesis in which the regression coefficient β_i is not zero. If the null hypothesis is rejected, the alternative hypothesis is adopted, and the explanatory variable X_i explains Y statistically significantly. For the hypothesis testing, we use *t*-statistic defined in Equation (5) to verify the statistical significance of keyword X_i with respect to Y (keyword *drone*) [27].

$$t_i = \frac{b_i - \beta_i}{\text{SE}(b_i)} \tag{5}$$

where b_i is the estimator of β_i computed by minimizing the sum of squares of the error in the model of Equation (3). $SE(b_i)$ is standard error (SE) of b_i corresponding to keyword X_i. We can reject or accept H_0 by the value of *t*-statistics of (5). The larger the value, the more statistically significant the explanatory power of the corresponding keyword for *drone*. We used the statistic to find the significance probability (*p*-value) and used this value to determine the statistical significance of each keyword. For example, at the 95% confidence level, we determine a keyword with *p*-value less than 0.05 as statistically significant, that is, we reject H_0 [28]. In this paper, we selected the explanatory variables (keywords) with *p*-value less than 0.05 in the multiple regression model.

Lastly, we extracted the important keywords for drone technology using BART modeling. Compared to the existing classification and regression tree models that rely on only ensemble methods, BART is based on Bayesian probability distribution as well as ensemble of trees [29]. Therefore, BART is an regression approach based on nonparametric Bayesian [30]. This finds unknown function *f* using recursive partitioning of predictor space. In this paper, the dimension of predictor space is *p* and the size of this dimension were equal to the number of explanatory variables (keywords). In our BART model, we define the fundamental equation as (6).

$$Y = \text{f}(X) + \varepsilon \tag{6}$$

The response variable Y is the keyword drone, and the explanatory variables are represented by $X = (X_1, X_2, \ldots, X_p) = (Keyword_1, Keyword_2, \ldots, Keyword_p)$. The error term ε is followed to normal distribution $N(0, \sigma^2)$. We make inference about model $f(\cdot)$. Using sum of *m* distinct regression trees, we expressed the model of Equation (6) as Equation (7) as follows [29–32].

$$Y = \text{Tree}_1^M(X) + \text{Tree}_2^M(X) + \cdots + \text{Tree}_m^M(X) + \varepsilon \tag{7}$$

where $Tree_i$ is *i*th regression tree. Moreover, *M* represents the parameters of terminal nodes as follows: $M = (\gamma_1, \gamma_2, \ldots, \gamma_r)$, where γ_i is *i*th parameter value of each node and *b* is the number of terminal nodes of each tree. Tree^M is entire tree with leaf parameters. Therefore, using *m* regression trees, we find the influence of each explanatory keyword *X* on response keyword Y (*drone*). In other words, using the results of relative influence of each explanatory keyword, we select important explanatory keywords that have a meaningful impact on the drone technology. Using the concept of Equation (7), the fundermental Equation (6) is represented as Equation (8).

$$\text{drone} = g(X; \text{Tree}, M) + \varepsilon, \ \varepsilon \sim N(0, \sigma^2) \tag{8}$$

Under Equation (8), the conditional mean of keyword drone given explanatory keyword X, $E(\text{drone}|X)$ is the terminal node parameter assigned by $g(X; \text{Tree}, M)$. Next, we impose a prior to all the parameters of the sum of regression trees. The prior distribution consists of tree structure, leaf parameters, and error variance (σ^2) in (9) [30].

$$P(\text{Tree}_1^M, \text{Tree}_2^M, \ldots, \text{Tree}_m^M, \sigma^2) = P(\sigma^2) \prod_{i=1}^m P(M_i|\text{Tree}_i) P(\text{Tree}_i) \tag{9}$$

where $P(\text{Tree}_i)$ influences on the location of nodes in tree structure and $P(M_i|\text{Tree}_i)$ describes the control of leaf parameters. The prior component of error variance $P(\sigma^2)$ is distributed to inverse-gamma distribution [30]. Moreover, the likelihood function specifies the responses of terminal nodes. We extract samples from the posterior distribution with response variable *y* in (10).

$$P(\text{Tree}_1^M, \text{Tree}_2^M, \ldots, \text{Tree}_m^M, \sigma^2 | y) \tag{10}$$

In our research, we applied Markov chain Monte Carlo (MCMC) to generate samples from the posterior distribution [33]. Therefore, we selected the important keywords representing drone technology using BART with prior, likelihood, posterior, and MCMC. The keyword selection process is performed by the number of covariates (keywords except *drone*) affecting keyword *drone*. The process uses the proportions of included variables for extracting important keywords [31]. The proportion of times in each keyword is calculated as total splitting rules over a splitting rule. Finally, we constructed technology scenarios of drone using BART modeling with correlation networks. Moreover, the scenarios were applied to MOT of drone such as R&D planning and new services development in drone technologies.

3.6. Summary of Proposed Methodology

Our research contains entire process of patent analysis from patent searching to development of technology scenarios. Furthermore, we built the patent document–keyword matrix using text mining techniques. This matrix is used for quantitative patent analysis using visualization and Bayesian data analysis. Finally, we generated technology scenarios for development of drone technology using the results of patent keyword analysis of drone. Figure 5 shows summary of our proposed methodology.
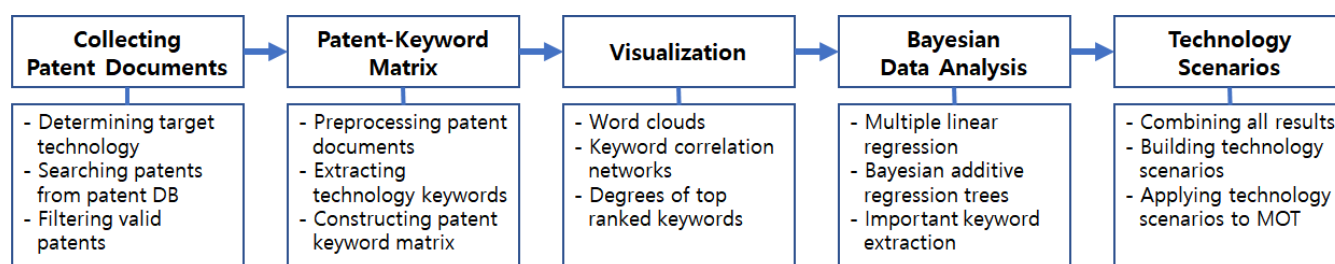
Figure 5. Correlation network structure.

In Figure 5, we illustrate the entirety of our proposed methodology described in Sections 3.1–3.5. Our final result, the technology scenarios, can be used in various fields of technology management, such as the establishment of R&D strategies. In next section, we collect and analyze drone patent documents as to how our proposed methodology can be applied to practical domains.

4. Experiments and Results

To show the validity and performance of proposed drone technology analysis, we collected and analyzed the patent documents related to drone technology. First, we retrieved the patent documents of drone technology from the popular patent databases, USPTO and WIPSON [13,14]. The searched patents include title, abstract, applicant, inventor, nation, international patent classification (IPC) code, claims, and citation information. There are many resources for describing technology in patent documents. In this paper, we selected the abstracts of issued patents. Of course, patent analysis on drone technology is possible using various features as well as abstracts. Our patent data analysis is focused on the keyword data extracted from the patent documents. The valid patent data contains 60,311 documents.

We performed our experiments in computational framework with 16 GB and Intel core i7 1.99 GHz for RAM and CPU, respectively. In addition, we used R data language and its packages for our patent data analysis [22–24,32]. First, we constructed the patent-keyword matrix and extracted 85 technology keywords from the matrix using the text-mining techniques. We used the papers, articles, and domain experts for extracting the keywords related to drone technology. Figure 6 shows the word cloud of drone technology keywords. We used this figure to find the keywords with high influence on drone technology. In the preprocessing process of text mining, we removed the suffixes of each keyword for the efficiency of structured data generation.
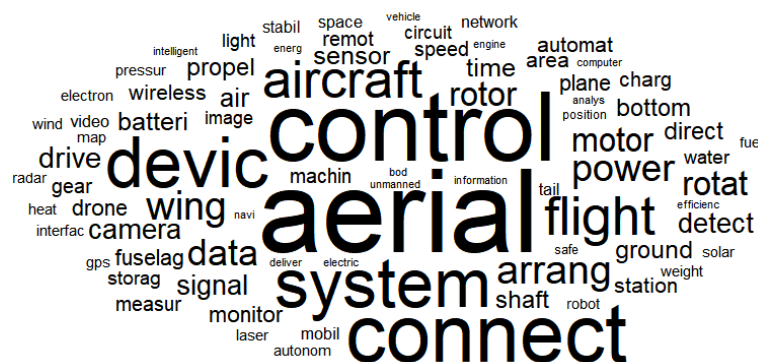

Figure 6. Word cloud of drone technology keywords.

For example, the keyword *devic* is the result of removing the suffix from device. From the result of word cloud, we found the keywords with high frequency such as *aerial, control, connect, devic, system, flight, aircraft, wing, arrang, power,* and *data*. They influence on the

development of drone technology. Table 3 illustrates the top 60 keywords representing the drone technology.

Table 3. Top 60 keywords of drone technology.

Ranking	Keywords
1~30	*aerial, control, connect, devic, system, flight, aircraft, wing, arrang, power, data, rotor, rotat, motor, drive, camera, signal, detect, air, time, batteri, sensor, ground, direct, bottom, shaft, propel, fuselag, monitor, drone*
31~60	*plane, automat, wireless, charg, area, station, remot, measur, gear, speed, machin, water, storag, image, light, circuit, tail, space, video, mobil, network, stabil, gps, interfac, weight, electron, pressur, wind, autonom, map*

Since it is difficult to obtain meaningful analysis results if the frequency of keywords is too small, we selected the top 60 keywords with high frequency as shown in Table 3. This was also shown in Figure 2. The first row shows the keywords with high ranked frequencies, and the second row illustrates the keywords with relatively low frequencies. Because keywords were extracted through the preprocessing of text mining, some keywords have suffixes omitted. Using the top 60 keywords, we constructed the keyword correlation networks according to the correlation coefficient. Figure 7 shows the keyword correlation network with correlation coefficient value larger than 0.2. In Figure 6, most keywords are isolated, but some keywords are connected each other. We found five connecting components in this figure. The keywords *motor*, *shaft*, and *rotat* were connected around *shaft*. We knew that the role of the shaft between the rotation of the drone propeller and the motor is important. In addition, the four keywords *flight*, *system*, *control*, and *remot* were connected around *control*. That is, the flight of drone and drone system are dependent on the remote control. The keywords *aerial* and *aircraft* are connected, and *ground* and *station* are also connected. Finally, *fuselag* and *tail* are connected via *wing*. The technology development of fuselage is based on the technologies of wing and tail of drone.
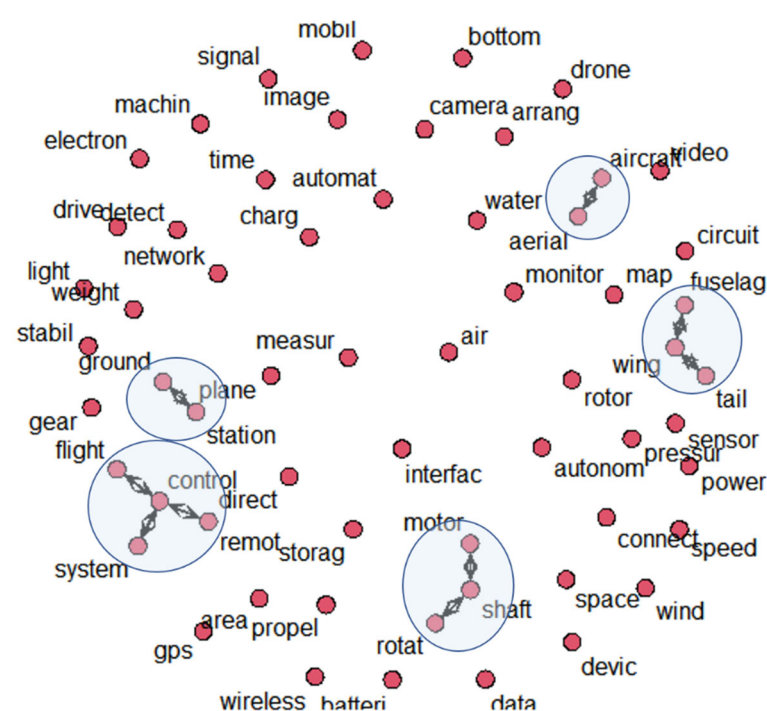

Figure 7. Keyword correlation network (correlation coefficient = 0.2).

Next, we built the correlation network with the value of correlation coefficient larger than 0.15 in Figure 8. We found six components from the network in Figure 8. The first component consisted of four keywords: *network*, *wireless*, *charg*, and *batteri*. This component represents the technology of battery charging in the environment of wireless networks. There were five keywords: *system*, *remot*, *flight*, *signal*, and *control* in the second component. The rest of the keywords were connected around keyword *control*. This component represents the technology of signal control for remote flight system. The third component contained the keywords *aerial* and *aircraft*. This means the technology for flying drones. The fourth component consisted of four keywords: *rotor*, *fuselage*, *tail*, and *wing*. The representative technology of this component is related to the fuselage of drones including rotor, wing, and tail. The fifth component is relatively large compared to others. This has eight keywords: *rotat*, *motor*, *drive*, *gear*, *shaft*, *connet*, *arrang*, and *bottom*. The component was composed of two keyword groups using keyword *shaft* as a medium. The first group consisting of *rotat*, *motor*, *drive,* and *gear* represented the technology related to drone driving. The second group with *connet*, *arrang*, and *bottom* illustrated the technology of drone arrangement. The keyword *shaft* denotes a technology for connecting two groups. The last component contained two keywords: *speed* and *electronic*. This component represents the technology related to speed of electronic signal.
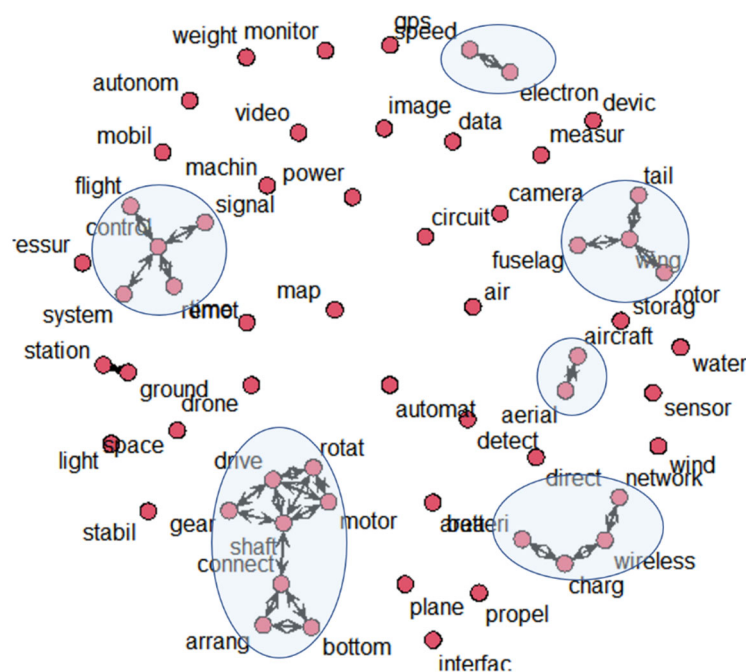

Figure 8. Keyword correlation network (correlation coefficient = 0.15).

In addition, we constructed the network by reducing the value of the correlation coefficient to 0.1 in order to understand the connection structure between more keywords. Figure 9 shows the correlation network with correlation coefficient of 0.1. We found three components in this network. The first component with the keywords *speed* and *electronic* was the same as the last component of the network with correlation coefficient of 0.15 in Figure 8. The second component included the keywords *pressur* and *air*, and therefore the representative technology of this component is the technology for air pressure. The third component had a large structure containing many keywords. In the networks of previous figures, various single components were merged into one component, and new keywords were added to this component. We identified additional detailed technologies required for developing a drone system through newly added keywords such as *data*, *sensor*, *video*, *camera*, *monitor*, and *automat*. We found that the technologies related to sensing, camera monitoring, and automation.
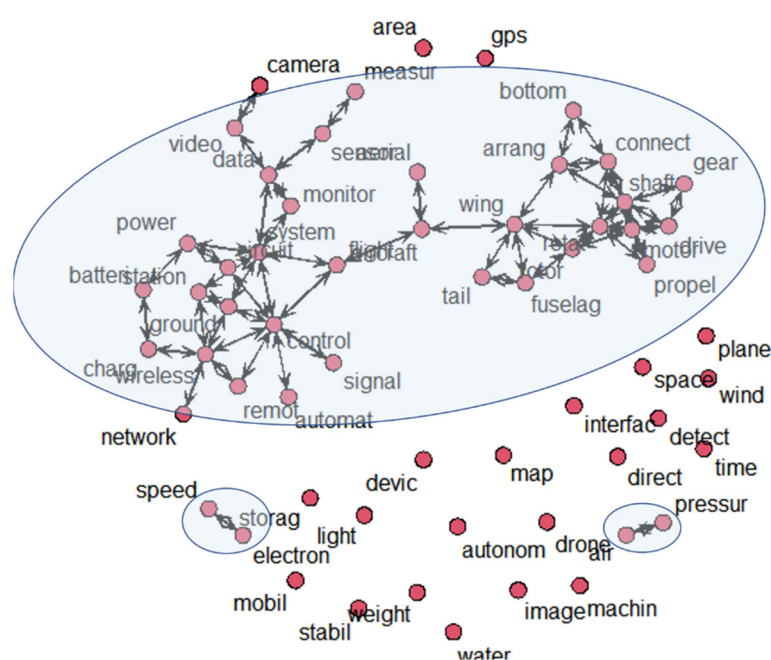
Figure 9. Keyword correlation network (correlation coefficient = 0.1).

We summarized the results of three keyword correlation networks to show the degrees of the keywords in Table 4. Each degree value was calculated as the sum of indegree and outdegree values by Equation (2). All connections between keywords in the correlation networks of Figures 6–8 are undirected (or bidirectional) networks. Therefore, in this paper, we provided a value of 2 to the connection of two keywords having indegree and outdegree at the same time as follows: degree value = indegrees + outdegrees. Table 4 shows the degree values of the top 60 keywords of drone technology.

Table 4. Degrees of top 60 keywords of drone technology.

Keyword	Correlation Coefficient			Keyword	Correlation Coefficient		
	0.2	0.15	0.1		0.2	0.15	0.1
aerial	2	2	2	plane	0	0	0
control	6	8	16	automat	0	0	2
connect	0	6	10	wireless	0	4	12
devic	0	0	0	charg	0	4	4
system	2	2	14	area	0	0	0
flight	2	2	6	station	2	2	6
aircraft	2	2	6	remot	2	2	4
wing	4	6	12	measur	0	0	2
arrang	0	4	8	gear	0	4	4
power	0	0	6	speed	0	2	2
data	0	0	8	machin	0	0	0
rotor	0	2	10	water	0	0	0
rotat	2	6	14	storag	0	0	0
motor	2	6	12	image	0	0	0
drive	0	8	8	light	0	0	0
camera	0	0	2	circuit	0	0	4
signal	0	2	2	tail	2	2	4
detect	0	0	0	space	0	0	0
air	0	0	2	video	0	0	4
time	0	0	0	mobil	0	0	0

batteri	0	2	4	network	0	2	2
sensor	0	0	4	stabil	0	0	0
ground	2	2	8	gps	0	0	0
direct	0	0	0	interfac	0	0	0
bottom	0	4	4	weight	0	0	0
shaft	4	10	16	electron	0	2	2
propel	0	0	6	pressur	0	0	2
fuselag	2	2	6	wind	0	0	0
monitor	0	0	4	autonom	0	0	0
drone	0	0	0	map	0	0	0

The larger the degree value of a keyword, the greater the influence of the keyword on drone technology. If degree value of a keyword is 0, it means that technology based on the keyword does not significantly affect drone technology. In Table 4, we also illustrate the degree values according to the values of correlation coefficient between keywords. As the value of the correlation coefficient decreases, the degree value increases because the number of connections between keywords increases. Therefore, for efficient comparison of degree values between keywords, we must also consider relative comparisons according to correlation coefficient values.

We also performed statistical significance tests of explanatory keywords for response keyword *drone* using multiple linear regression model. Table 5 shows the testing results of the explanatory keywords. In the table, we represented *t*-statistic and *p*-value of parameter according to each explanatory keyword for response keyword *drone*. At the 95% confidence level, the keywords with *p*-value less than 0.05 are statistically significant. Therefore, we selected the keywords *aerial, devic, system, wing, data, rotat, motor, drive, camera, signal, time, batteri, shaft, fuselag, plane, automat, area, station, gear, water, image, light, circuit, tail, space, network, stabil, pressur,* and *autonom*. We can use these keywords to describe the core technologies required for drone technology development.

Table 5. Testing results of drone keywords in linear model (response = drone).

Explanatory	*t*	*p*-Value	Explanatory	*t*	*p*-Value
(Intercept)	43.988	0.0001	plane	−7.298	0.0001
aerial	−26.258	0.0001	automat	−2.119	0.0341
control	−1.144	0.2527	wireless	−0.921	0.357
connect	−10.456	0.0001	charg	1.945	0.0518
devic	−0.285	0.7759	area	4.639	0.0001
system	1.821	0.0686	station	11.029	0.0001
flight	−0.587	0.5574	remot	1.618	0.1057
aircraft	−17.125	0.0001	measur	−1.655	0.0979
wing	−0.643	0.5205	gear	−2.839	0.0045
arrang	−4.414	0.0001	speed	−1.251	0.211
power	−2.567	0.0103	machin	−0.952	0.3413
data	3.203	0.0014	water	−2.597	0.0094
rotor	−4.426	0.0001	storag	−0.458	0.6468
rotat	−2.672	0.0075	image	−9.271	0.0001
motor	−2.273	0.0231	light	−2.157	0.031
drive	−2.837	0.0045	circuit	−3.514	0.0004
camera	1.667	0.0956	tail	−2.01	0.0445
signal	3.379	0.0007	space	−1.985	0.0472
detect	−0.994	0.3202	video	−0.67	0.5025
air	−4.641	0.0001	mobil	0.131	0.8956
time	−6.071	0.0001	network	6.67	0.0001

batteri	−3.209	0.0013	stabil	−2.956	0.0031
sensor	2.424	0.0153	gps	−1.282	0.1997
ground	−6.253	0.0001	interfac	0.715	0.4748
direct	2.631	0.0085	weight	−0.538	0.5909
bottom	−4.616	0.0001	electron	0.685	0.4931
shaft	−1.3	0.1937	pressur	−3.873	0.0001
propel	−0.234	0.8151	wind	−0.067	0.9462
fuselag	−4.189	0.0001	autonom	7.722	0.0001
monitor	−1.254	0.21	map	−1.416	0.1569

The larger the absolute *t*-statistic value of a keyword, the more it means that the keyword has more influence on drone technology development. Moreover, the sign of the value indicates the direction of technological influence. For example, if the sign of the value becomes negative, it exerts an influence in the opposite direction. Using the results of multiple regression, we carried out the patent keyword analysis using BART modeling. First, we had to select the number of trees in BART model. Figure 10 shows the root mean squared error (RMSE) by number of trees. We found that the RMSE value with number of trees of 30 was the lowest in Figure 10. Therefore, we determined the number of trees to be 30 for BART modeling.
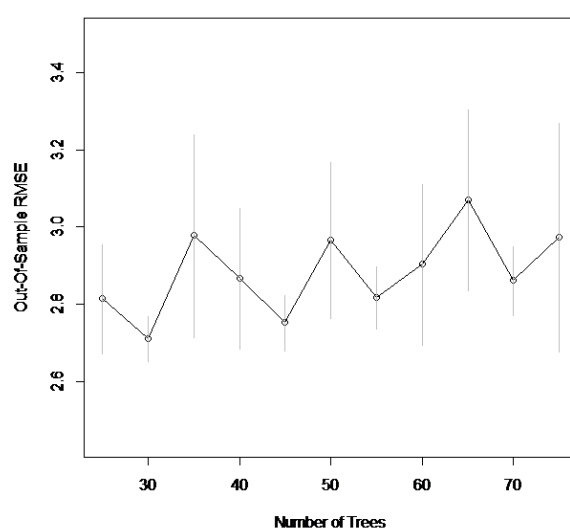

Figure 10. RMSE for prediction performance by number of trees.

Before analyzing patent data analysis using BART model, we performed Q-Q plotting to check whether the data were satisfied with the normality assumption. Figure 11 illustrates the Q-Q plot of the patent keyword data.
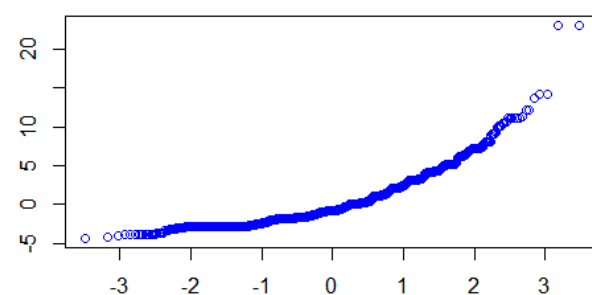

Figure 11. Q–Q plot of patent keyword data for BART modeling.

We can visually check whether the data satisfies the normality assumption through the Q-Q plot results. That is, we identify how close the data are to normal distribution in Figure 10. The *y*-axis of the plot represents z-scores of data, and the *x*-axis is the corresponding quantile of normal distribution. Figure 10 shows the data points roughly are out of the diagonal line. Therefore, we carried out additional normality tests such as the Shapiro–Wilk and Kolmogorov–Smirnov tests [34]. The test results are shown in Table 6.

Table 6. Normality test results.

Test Method	*p*-Value
Shapiro–Wilk	<0.0001
Kolmogorov–Smirnov	<0.0001

As shown in Table 6, we performed two statistical tests to confirm whether the patent data used in this paper satisfied the normality assumption. All *p*-values of Shapiro–Wilk and Kolmogorov–Smirnov tests were less than 0.0001. Thus, we found that the data were not satisfied with the normality assumption. This means that when data analysis was conducted from the point of view of frequentist using only the likelihood function of given data, we encountered the problem of performance degradation of analysis results. To overcome this problem, we needed the Bayesian approach, which uses not only likelihood but also prior.

In general, most Bayesian modeling requires MCMC computing for sampling from posterior distribution. We also used the MCMC method in our BART modeling. In MCMC, the time it takes to converge to the posterior distribution is called the burn-in [33]. The samples generated before the burn-in are not used for BART modeling. This is because most of the samples extracted at the initial time deviate from the posterior distribution. Figure 12 shows the burn-in iterations in our MCMC computing. The *x*-axis and *y*-axis are number of MCMC iteration and trees acceptance, respectively.
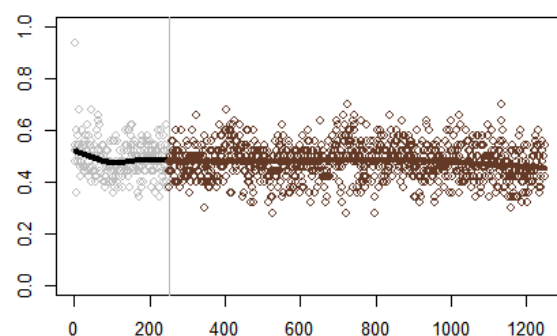

Figure 12. Burn-in iteration for MCMC computing.

From the result of Figure 11, we found that the number of iterations for burn-in was about 250. Thus, we determined the burn-in time to be 250 in our experiments. Next, we visualized the result of important keyword selection of BART modeling, as shown in Figure 13. This plot shows the relative importance of explanatory keywords that affect drone technology.
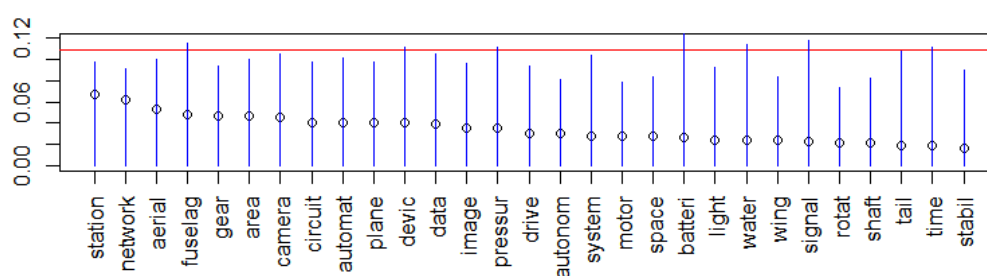
Figure 13. Variable importance results of explanatory keywords.

In Figure 13, our BART model selected 29 keywords that affect drone technology out of the 60 keywords. We can see that keyword *station* had the most influence on drone technology. We confirmed that the keywords affect drone technology in the order of *network*, *aerial*, *fuselag* (fuselage), etc., after *station*. In this paper, the keyword drone was used for response variable, and the others except drone were used for explanatory variables. In our work, we can select the significantly explanatory keywords for response keyword by various methods. We first performed the keyword visualization on the basis of word cloud, correlation networks with degrees to find the keywords that affect drone. Moreover, we carried out multiple liner regression and BART models to select the significant keywords to drone. Finally, we created the technology scenarios in Table 7 on the basis of all the results obtained in this paper, including the results of 29 keywords selected in Figure 12.

Table 7. Technology scenarios for drones.

Scenario	Technology Description	Keyword
First	Technology to improve flight control and safety of the fuselage by collecting and analyzing visual information around the drone	*aircraft, automat, autonom, camera, data, direct, drive, fuselag, gear, image, network, plane, rotat, sensor, shaft, signal, space, stabil, tail, wing*
Second	Technology that collects surrounding information based on sensor signals and uses it to generate and control power for flight	*aerial, area, arrang, circuit, control, data, detect, flight, fuselag, machin, measur, monitor, motor, power, pressur, propel, remot, sensor, signal, speed*
Third	Technology to control efficient battery storage and operation for drone flight	*aerial, aircraft, batteri, bottom, charg, control, electron, flight, ground, interfac, motor, power, pressur, speed, station, storag, system, water, weight, wind*

We show technology scenarios with first, second, and third ranked technologies to affect the development of drone technology. The first scenario is defined as the technology to improve flight control and safety of the fuselage by collecting and analyzing visual information around the drone. This technology is based on 20 keywords as follows: *aircraft, automat, autonom, camera, data, direct, drive, fuselag, gear, image, network, plane, rotat, sensor, shaft, signal, space, stabil, tail,* and *wing*. The second core technology for drone is the technology that collects surrounding information on the basis of sensor signals and uses it to generate and control power for flight. The keywords of *aerial, area, arrang, circuit, control, data, detect, flight, fuselag, machin, measur, monitor, motor, power, pressur, propel, remot, sensor, signal,* and *speed* are composed of this technology. Each keyword becomes sub technology that constitutes the core technologies of drones. The last is the technology to control efficient battery storage and operation for drone flight on the basis of the 20 keywords of *aerial, aircraft, batteri, bottom, charg, control, electron, flight, ground, interfac, motor, power, pressur, speed, station, storag, system, water, weight,* and *wind*. According to the experts in

drone industry, the visualization and BART results of this paper can be used to create a broader range of technology scenarios for drone technology development. In our case study, we did not compare the model accuracy or computation time of our proposed method with other machine learning methods. This was because the purpose of our research was to find the core technologies required for drone development and use these results to create technology scenarios necessary for MOT of drones. Furthermore, our proposed method of drone technology analysis can be expanded to other technology domains such as artificial intelligence, big data, and internet of things.

5. Discussion

The uses of ensemble learning in statistics and machine learning have been applied to various tasks for learning from data. The ensemble learning is to build various models and combine the analysis results to get improved performance [20]. This is an approach opposite to a single model that uses only one analysis result. BART is an analytic model based on ensemble for trees. In this paper, we tried to improve the performance of patent analysis using this model. The BART used for our study combines a set of tree models and identifies which explanatory keywords are more important for explaining the response keyword drone. This is based on the Bayesian inference model with nonparametric function estimation using regression trees [30,32]. The BART has prior with tree structure and leaf parameters, given that tree structure and error variance are independent of tree structure and leaf parameters [30]. The prior is combined with likelihood to make posterior, and the posterior is able to build a model with better performance than the other models such as gradient boosting tree [35,36]. Therefore, we used the BART that relies on the Bayesian probability model to obtain better performance than others.

Next, we presented the implications of our research results from two perspectives. The first is the contribution of our research from a practical point of view. Our proposed method will contribute to the technology management of a company. Companies will use the results of technology analysis to establish their own R&D strategies and develop new products. Therefore, using the results of technology analysis by our proposed methodology, the company can lead the market with technological competitiveness. Another contribution of our research is the development of technology analysis methodology from an academic point of view. We have constantly developed various methodologies of technology analyzes. In most previous research related to technology analysis, the qualitative analysis that relied on the subjective knowledge and judgment of expert group was mainstream. However, recently, the quantitative methodologies that analyze patent data using statistics and machine learning algorithms are being actively developed. In this paper, we also proposed a quantitative method to analyze patent document data. We applied new visualization and advanced statistical method for analyzing patent keyword data. With the spread of artificial intelligence and big data, the importance of technology increases even more. Therefore, theoretical research and practical application of quantitative technology analysis using statistics and machine learning are continuously needed.

6. Conclusions, Limitations, and Future Research

In this paper, we proposed a methodology of patent analysis using keyword visualizations and BART modeling for technology analysis. We selected the drone technology as our target technology. Using word cloud, correlation networks, multiple linear regression, and BART modeling, we built technology scenarios for drone technology development. We constructed three scenarios as representative strategies for developing drone technology and service. The first scenario is the technology to control drone flighting by collecting and analyzing visual data around the drones. We generated this scenario through the technological relations between 20 keywords identified in the results of keyword visualization and BART modeling. We made the second scenario the technology to control the power of drone flighting by collecting surrounding information and using it. Lastly, we built the technological strategy to control efficient battery storage and

operation for drone flighting as the third scenario. All scenarios were produced by the analysis results on the basis of the keywords extracted from drone patents. The results of our research can be applied to various fields in MOT for drone technology. Through collaboration with a group of experts in the field of drone technology, we can draw more diverse implications of the research results of this paper. For example, the drone experts can create new technology scenarios using our results. We leave the further applications of this research results to the role of experts in the drone industry.

We analyzed the drone technology only using keywords extracted from drone patent documents. However, in addition to patent keywords, much more information describing the technology, such as application date, citation information, pictures, and technology classification codes, is contained in the patent document. Therefore, we can expect better analysis results if we perform technology analysis using more information from patent documents. This study has a limitation in that we performed technology analysis using only a portion of the large patent data.

In our research, we developed keyword visualization using word cloud and correlation networks. We also constructed a patent analysis model based on multiple linear regression and BART. There are many methods for patent data analysis from statistics and machine learning. To construct more advanced models that analyze patent data, we will consider diverse learning algorithms based on deep learning and mathematical statistics. Therefore, in our future research, we will develop a better patent analysis methodology than now through two new trials. First, we will extract more information contained in patent documents as well as technology keywords and use them for analysis. Second, we will develop a new learning model for analysis optimized for patent data. Through this, we will continue to develop new and excellent technology analysis models for technology management.

Author Contributions: S.P. designed this research and collected the data set for the experiment. S.P. and S.J. analyzed the data to show the validity of this paper, wrote the paper, and performed all the research steps. In addition, all authors have cooperated with each other for revising the paper. All authors have read and agreed to the published version of the manuscript.

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

References

1. Nouacer, R.; Hussein, M.; Espinoza, H.; Ouhammou, Y.; Ladeira, M.; Castiñeira, R. Towards a framework of key technologies for drones. *Microprocess. Microsyst.* **2020**, *77*, 103142.
2. Chen, J.; Wang, G.; Luo, L.; Gong, W.; Cheng, Z. Building Area Estimation in Drone Aerial Images Based on Mask R-CNN. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 891–894.
3. Holden, P.B.; Rebelo, A.J.; New, M.G. Mapping invasive alien trees in water towers: A combined approach using satellite data fusion, drone technology and expert engagement. *Remote Sens. Appl. Soc. Environ.* **2021**, *21*, 100448.
4. Kornatowski, P.M.; Mintchev, S.; Floreano, D. An origami-inspired cargo drone. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Vancouver, BC, Canada, 24–28 September 2017; pp. 6855–6862.
5. Suzuki, S. Recent researches on innovative drone technologies in robotics field. *Adv. Robot.* **2018**, *32*, 1008–1022.
6. Tejado-Ramos, J.; Chocarro-León, M.; Barrero-Béjar, I.; Valverde-Calvo, A.; Ferreras-Moreno, M.; Giraldo-Pavón, F.; Tarragona-Pérez, C. Enhancement of the sustainability of wolfram mining using drone remote sensing technology. *Remote Sens. Appl. Soc. Environ.* **2021**, *23*, 100542.
7. Tzelepi, M.; Tefas, A. Graph Embedded Convolutional Neural Networks in Human Crowd Detection for Drone Flight Safety. *IEEE Trans. Emerg. Top. Comput. Intell.* **2021**, *5*, 191–204.
8. Kwon, H.; Kim, J.; Park, Y. Applying LSA text mining technique in envisioning social impacts of emerging technologies: The case of drone technology. *Technovation* **2017**, *60–61*, 15–28.
9. Hunt, D.; Nguyen, L.; Rodgers, M. *Patent Searching Tools & Techniques*; Wiley: Hoboken, NJ, USA, 2007.
10. 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.
11. Feng, X.; Fuhai, L. Patent Text Mining and Informetric-based Patent Technology Morphological Analysis: An Empirical Study. *Technol. Anal. Strateg. Manag.* **2021**, *24*, 467–479.

12. Uhm, D.; Ryu, J.; Jun, S. Patent Data Analysis of Artificial Intelligence Using Bayesian Interval Estimation. *Appl. Sci.* **2020**, *10*, 570.
13. USPTO. The United States Patent and Trademark Office. Available online: http://www.uspto.gov (accessed on 15 March 2021).
14. WIPSON. WIPS Corporation. Available online: http://www.wipson.com (accessed on 1 April 2021).
15. KIPRIS. Korea Intellectual Property Rights Information Service. Available online: www.kipris.or.kr (accessed on 1 April 2021).
16. Berman, J.J. *Principles and Practice of Big Data: Preparing, Sharing, and Analyzing Complex Information*, 2nd ed.; Academic Press: London, UK, 2018.
17. Ryu, J.; Jun, S. A Superpopulation Model for Patent Big Data Analysis. *Int. J. Softw. Eng. Its Appl.* **2016**, *10*, 153–162.
18. Park, S.; Lee, S.; Jun, S. Patent Big Data Analysis using Fuzzy Learning. *Int. J. Fuzzy Syst.* **2017**, *19*, 1158–1167.
19. Chakraborty, M.; Byshkin, M.; Crestani, F. Patent citation network analysis: A perspective from descriptive statistics and ERGMs. *PLoS ONE* **2020**, *15*, e0241797.
20. Bruce, P.; Bruce, A.; Gedeck, P. *Practical Statistics for Data Scientists*; O'Reilly Media: Sebastopol, CA, USA, 2020.
21. Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Technique*s, 3rd ed.; Morgan Kaufmann: Waltham, MA, USA, 2012.
22. Feinerer, I.; Hornik, K.; Meyer, D. Text mining infrastructure in R. *J. Stat. Softw.* **2008**, *25*, 1–54.
23. Feinerer, I.; Hornik, K. *Package 'tm' Version 0.7-8, Text Mining Package*; CRAN of R Project; R Foundation for Statistical Computing: Vienna, Austria, 2020.
24. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria. Available online: http://www.R-project.org (accessed on 1 April 2021).
25. Abazi-Bexheti, L.; Kadriu, A.; Apostolova, M. Word Cloud Analytics of the Computer Science Research Publications' Titles over the Past Half Century. In Proceedings of the 43rd International Convention on Information, Communication and Electronic Technology, Opatija, Croatia, 28 September–2 October 2020; pp. 887–892.
26. Heimerl, F.; Lohmann, S.; Lange, S.; Ertl, T. Word Cloud Explorer: Text Analytics Based on Word Clouds. In Proceedings of the 47th Hawaii International Conference on System Sciences, Waikoloa, HI, USA, 6–9 January 2014; pp. 1833–1842.
27. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
28. Hogg, R.V.; Tanis, E.A.; Zimmerman, D.L. *Probability and Statistical Inference*, 9th ed.; Pearson: Essex, UK, 2015.
29. Chipman, H.A.; George, E.I.; McCulloch, R.E. BART: Bayesian Additive Regression Trees. *Ann. Appl. Stat.* **2010**, *4*, 266–298.
30. Kapelner, A.; Bleich, J. bartMachine: Machine Learning with Bayesian Additive Regression Trees. *J. Stat. Softw.* **2016**, *70*, 1–40.
31. Bleich, J.; Kapelner, A.; Jensen, S.T.; George, E.I. Variable Selection for BART: An Application to Gene Regulation. *Ann. Appl. Stat.* **2014**, *8*, 1750–1781.
32. Sparapani, R.; Spanbauer, C.; McCulloch, R. Nonparametric Machine Learning and Efficient Computation with Bayesian Additive Regression Trees: The BART R Package. *J. Stat. Softw.* **2021**, *97*, 1–66.
33. Gelman, A.; Carlin, J.B.; Stern, H.S.; Dunson, D.B.; Vehtari, A.; Rubin, D.B. *Bayesian Data Analysis*, 3rd ed.; Chapman & Hall/CRC Press: Boca Raton, FL, USA, 2013.
34. Rice, J.A. *Mathematical Statistics and Data Analysis*, 2nd ed.; Duxbury Press: Belmont, CA, USA, 1995.
35. Du, Q.; Li, N.; Yang, S.; Sun, D.; Liu, W. Integrating KNN and Gradient Boosting Decision Tree for Recommendation. In Proceedings of the IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference, Chongqing, China, 12–14 March 2021; pp. 2042–2049.
36. Wu, W.; Xia, Y.; Jin, W. Predicting Bus Passenger Flow and Prioritizing Influential Factors Using Multi-Source Data: Scaled Stacking Gradient Boosting Decision Trees. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 2510–2523.