

Article

Ensemble Modeling for Sustainable Technology Transfer

Junseok Lee ¹, Ji-Ho Kang ¹, Sunghae Jun ², Hyunwoong Lim ¹, Dongsik Jang ¹ and Sangsung Park ^{3,*}

¹ Department of Industrial Management Engineering, Korea University, Seoul 02841, Korea; jxli12@korea.ac.kr (J.L.); kangmae@korea.ac.kr (J.-H.K.); limhw@korea.ac.kr (H.L.); jang@korea.ac.kr (D.J.)

² Department of Big Data and Statistics, Cheongju University, Chungbuk 28503, Korea; shjun@cju.ac.kr

³ Graduate School of Management of Technology, Korea University, Seoul 02841, Korea

* Correspondence: hanyul@korea.ac.kr; Tel.: +82-2-3290-4618

Received: 19 May 2018; Accepted: 29 June 2018; Published: 2 July 2018



Abstract: These days, technological advances are being made through technological conversion. Following this trend, companies need to adapt and secure their own sustainable technological strategies. Technology transfer is one such strategy. This method is especially effective in coping with recent technological developments. In addition, universities and research institutes are able to secure new research opportunities through technology transfer. The aim of our study is to provide a technology transfer prediction model for the sustainable growth of companies. In the proposed method, we first collected patent data from a Korean patent information service provider. Next, we used latent Dirichlet allocation, which is a topic modeling method used to identify the technical field of the collected patents. Quantitative indicators on the patent data were also extracted. Finally, we used the variables that we obtained to create a technology transfer prediction model using the AdaBoost algorithm. The model was found to have sufficient classification performance. It is expected that the proposed model will enable universities and research institutes to secure new technology development opportunities more efficiently. In addition, companies using this model can maintain sustainable growth in line, coping with the changing pace of society.

Keywords: technology transfer; prediction model; latent Dirichlet allocation; technology topic; ensemble model

1. Introduction

Technology is a very important tool in the today's information society. Especially in recent years, as innovation is emphasized, the importance of technology is being recognized again. Technological innovation consists of the creation and application of knowledge, and it is an important factor in sustainable growth [1,2]. In the past, companies operated independently in the areas of research and development, management, and marketing. As mentioned above, however, as the importance of technology is emphasized, companies are pursuing the concept of technology management (TM), which considers technology and management together. Many companies have tried to standardize technologies or manage intellectual properties in order to secure a competitive advantage through TM [1].

Intellectual property management (IPM) is a managerial method that strategically utilizes intellectual property, such as patents, trademarks, and copyrights, for corporate management [3]. Among these, a patent discloses information about an invention through the law, and the inventor is able to achieve exclusive rights to the invention. In other words, although a patent possesses a disadvantage in terms of its potential to be disclosed to the public, the inventor can have exclusive rights to the invention for 20 years. For this reason, many technology-based companies try to acquire exclusive

rights through patent applications after completing the research and development (R&D) of a technology. In this regard, patents are considered very useful by helping not only to preserve technology rights, but also to identify the direction of R&D and the competition intensity in the market [4]. Researchers who consider the characteristics of a patent, as previously mentioned, have attempted to analyze patents from a macro perspective and to forecast the direction of technology innovation [5–9].

A patent is recognized as a means of creating profit through licensing or technology transfer, as well as protecting the rights to an invention. This can be explained through the following example. Qualcomm has become a strong player in the telecom market with its huge royalty income from its Code Division Multiple Access (CDMA) technology. Another example is the recently emerged patent trolls, formally referred to as non-practicing entities (NPEs) [10,11]. In particular, a patent troll does not develop the actual invention but is allowed to purchase a patent from patent holders and retain it. If another company infringes on the patent it purchased, it will try to gain economic advantage through litigation [10]. As such, a patent has economic advantages, as well as legal rights.

Corporations, research institutes, and universities put forth a great deal of effort to transfer their technologies in order to create economic value from the patents. Technology transfer (TT) means that the holder of the knowledge, know-how, or technical materials transfers ownership or rights to the technologies to others who require them [12,13]. From the perspective of technology holders, it is advantageous because they can achieve financial profit to invest in a new research and development task when they sell their licensing or technical rights to another. On the other hand, from a buyers' perspective, it is beneficial because they can reduce the time and cost required for the technology development process. In recent years, we can see that innovation has been taking place through technology convergence [6,14,15]. From this perspective, TT is a good way to cope with the patterns of technological change. Furthermore, many scholars are doing a lot of research to promote TT that has a beneficial impact on both innovation and economy. Lai and Tsai introduced an evaluation methodology to enable efficient technology transfer. The proposed method is based on the analytic hierarchy process (AHP) and fuzzy logic method. They analyzed 20 questionnaires on Taiwan's machinery industry to demonstrate the effectiveness of the proposed method. Although this article suggested a methodology of using scientific methods to evaluate the effectiveness of TT, the study is limited because only survey data was used [16]. In order to help technology developers plan their R&D process in a multi-technology industry, Park et al. proposed a method of analyzing the possibility of technology transfer using patent citation information [17,18]. In addition, Choi et al. showed a technology prediction model based on the scientific method that uses patent information [13]. They extracted the quantitative indexes included in the patents for each organization through patent analysis and proposed an integrated technology transfer prediction model using a regression model, social network analysis, and decision tree algorithm.

When evaluating the quality of a patent, it is very important to consider the technological description, as well as qualitative factors. Nevertheless, previous studies have only considered quantitative indicators or technical material. The goal of this study is to propose a technology transfer prediction model based on the AdaBoost algorithm that uses technology topics and quantitative indexes of a patent to solve the above limitation. This paper is organized as follows. In Section 2, we describe the core theories used by the proposed model, and we propose a technology transfer prediction model using the ensemble method in Section 3. To illustrate the proposed model, the authors carry out a case study using the proposed methodology in Section 4. In Section 5, we discuss the conclusion of our experiment and suggest areas for future study at the end.

2. Background

This paper presents a novel methodology for predicting technology transfer, which considers both the technology description and quantitative indexes of a patent. The prediction model proposed in this study is based on the following theoretical backgrounds: patent analysis, topic model, and the AdaBoost algorithm. This section explains each of these.

2.1. Patent Analysis for Sustainable Technology Management

A patent is an essential element for protecting the rights of inventions and securing technological competitiveness. It is a necessary tool for technology management and can also be used for technology forecasting and R&D management [4,7,9,14,19–21]. In the initial research on technology management using patent data, Biju et al. used only quantitative information from five-year patent data from India to identify trends and innovation levels in India [5]. Their study found macroscopic trends and levels of technology, but it is limited because of the difficulty of understanding the details. On the other hand, Yoon and Park analyzed patent data using a text mining technique [19,20]. They extracted keywords from the patent information literature by text mining and analyzed it using a morphological approach. Through this, they attempted to discover technological opportunities, as well as perform technology forecasting. Daim et al. forecast emerging technology by analyzing the bibliometrics of patents [7]. Jun and Park suggested a comprehensive patent analysis method that considered time-series information, information from literature, and international patent classification (IPC), which is classified according to the technical description of a patent by an examiner. They evaluated the level of technological innovation using the proposed methodology [6,8]. Recently, patent analysis methodology based on statistical techniques has been proposed for technology forecasting and discovering emerging technology. Kim and Jun (2016) focused on the fact that the frequency of words occurring in patent documents is mostly zero. They attempted a technical analysis using a zero-inflated Poisson distribution and a negative binomial regression [22]. In addition, Park and Jun found that the frequency of a keyword and IPC generated in the patent documents follows the Poisson distribution. Considering this, they have constructed a technical analysis model for the LED field through a Poisson regression model and Bayesian network [15]. On the other hand, many studies have been carried out to discover promising technologies and emerging technologies through patterns generated in patent applications using machine learning. Kim and Bae attempted to foresee promising technology by identifying patterns of keywords in patent documents using clustering, a typical unsupervised technique [9]. Kyebambe et al. try to explore emerging technology utilizing supervised learning that uses patent citation information [21].

2.2. Topic Modeling

The topic model is one of the unsupervised methods; that is, it is a text mining technique with which the topics or themes of documents can be identified from a larger collected document corpus [23,24]. Latent Dirichlet allocation (LDA), which is one of the most popular topic modeling techniques, is a probabilistic model for expressing a corpus based on Bayesian models and is also considered to be a probabilistic extension of latent semantic analysis (LSA) [23,24]. The basic idea of the LDA is that each document has a topic, and a topic can be defined as a word distribution. Figure 1 shows a graphical model of LDA.

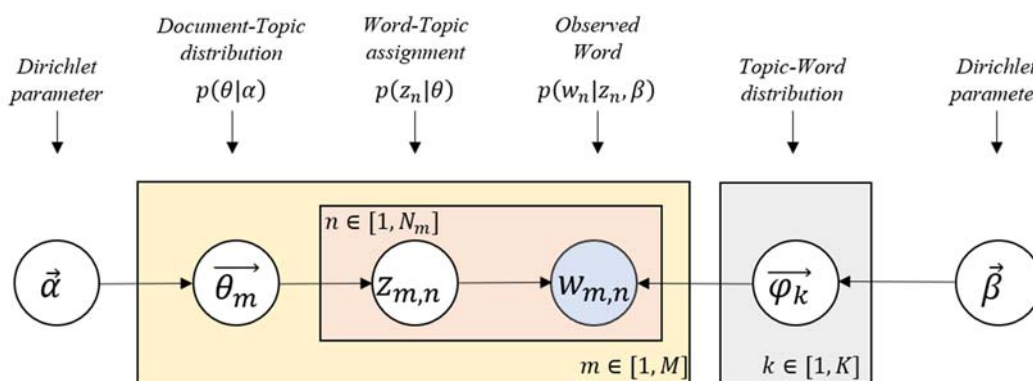


Figure 1. Graphical model of latent Dirichlet allocation (LDA).

When there is an m -th document D_m , the distribution in which the document D_m is included in the latent topic $z_{m,n}$ is denoted by $\vec{\theta}_m$, which is a multinomial distribution whose hyper-parameter α follows the Dirichlet distribution. The number of topics $k = 1, \dots, K$, can be statistically estimated, or the experimenter can determine a fixed value. The distribution of words for k topics is denoted by $\vec{\varphi}_m$, which is also a multinomial distribution whose hyper-parameter β follows the Dirichlet distribution. The word probability $w_{m,n}$ is determined by $p(w_{m,n}|z_{m,n}, \beta)$.

$$p(\theta, z, w|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^N p(z_n|\theta) p(w_n|z_n, \beta) \tag{1}$$

In Equation (1), the marginal distribution of a document can be obtained as shown in Equation (2) integrating over θ and summing over z .

$$p(w|\alpha, \beta) = \int p(\theta|\alpha) \left(\prod_{n=1}^N \sum_{z_n} p(z_n|\theta) p(w_n|z_n, \beta) \right) d\theta \tag{2}$$

Finally, the probability of the corpus can be expressed as the product of the marginal probabilities of each single document, as shown in Equation (3).

$$p(D|\alpha, \beta) = \prod_{m=1}^M \int p(\theta_m|\alpha) \left(\prod_{n=1}^{N_m} \sum_{z_n} p(z_{m,n}|\theta_m) p(w_{m,n}|z_{m,n}, \beta) \right) d\theta_m \tag{3}$$

A topic model is a useful method to identify topics hidden in a set of documents. As a result of the advantage described above, in recent years, research is often carried out to analyze technical documents using topic models. Kim et al. used a topic model to find common technologies included in patent data [25]. Lee and Kang conducted a topic model analysis for published articles and attempted to discover critical topics in technology innovation management [26].

2.3. Ensemble Method—AdaBoost Algorithm

The ensemble method is an algorithm to improve accuracy by combining multiple weak learners, typically by boosting and bagging. The Adaboost algorithm is a method proposed by Y. Freund and R. Schapire in 1995 to improve ‘boosting’ [27,28]. The most important feature of the AdaBoost algorithm is that the weak learners generated in bagging are made in parallel, whereas AdaBoost produces sequentially weaker learners. To solve the multi-label case problem and improve the performance AdaBoost generalization, Schapire and Singer suggested a method for tuning the weight in 1999 [29]. As we consider the problem with a single label in our study, the method proposed by Freund and Schapire is used. Figure 2 shows a graphical model of the AdaBoost algorithm.

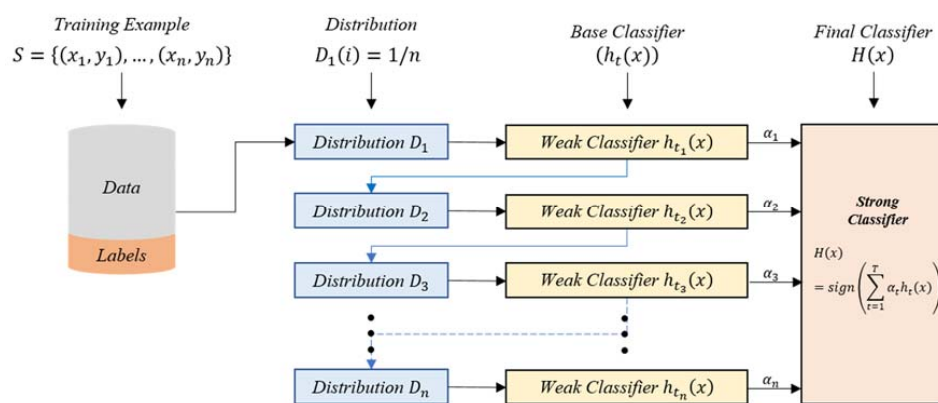


Figure 2. Graphical model of AdaBoost algorithm.

The main idea is to adjust the weights of weak classifiers for each distribution and finally combine these weak classifiers to get one strong classifier. In other words, it generates a hypothesis with high accuracy.

Suppose that the training example is given as $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, where $x_i \in \mathcal{X}$, $y_i \in \{-1, +1\}$, and y_i refer to class or label. An initial distribution $D_1(i)$ is defined as $D_1(i) = 1/n$, a uniform distribution. The distribution D_t is updated according to the weighted error ϵ_t to minimize the weak hypothesis h_t in each round.

$$\epsilon_t = Pr_{i \sim D_t}[h_t(x_i) \neq y_i] \quad (4)$$

According to error ϵ_t , weight α_t of weak hypothesis h_t is determined according to Equation (4), and the distribution $D_{t+1}(i)$ is updated through Equation (5) by the weight α_t .

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - \epsilon_t}{\epsilon_t} \right) \quad (5)$$

As can be seen from Equation (5), if $\epsilon_t > 0.5$, $\alpha_t < 0$ and if $\epsilon_t < 0.5$, $\alpha_t > 0$. That is, as the error ϵ_t of the weak hypothesis h_t decreases, the weight α_t increases. If the error $\epsilon_t > 0.5$, the performance of weak classifier is lower than that of random guessing. Therefore, the hypothesis of the weak classifier is not considered, and we move to the next round.

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \exp(-\alpha_t y_i h_t(i)) \quad (6)$$

In Equation (6), if $h_t(i) = y_i$, then $D_{t+1}(i) = \frac{D_t(i) \times e^{-\alpha_t}}{Z_t}$, and if $h_t(i) \neq y_i$, then $D_{t+1}(i) = \frac{D_t(i) \times e^{\alpha_t}}{Z_t}$, where Z_t represents a normalization factor. This satisfies $\sum_{i=1}^T D_t(i) = 1$. Finally, the strong classifier is defined as follows.

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right) \quad (7)$$

3. Methodology

In this research, we propose a scientific-based technology transfer prediction model using the AdaBoost algorithm for sustainable technology management. The proposed model is able to cover the quantitative indexes and information in the patent literature. Patent data are composed of various information on the inventions, such as text, numeric information, equations, and figures. In order to advance this experiment, therefore, patent data need to be changed into structured data. This study progresses as follows. In the first step, we identify the technology topics by analyzing titles and abstracts in patent data. In the next step, we extract quantitative indexes from patent data. Finally, merging the technology topics and quantitative indexes, the technology transfer prediction model is produced using AdaBoost. Figure 3 illustrates this experimental process.

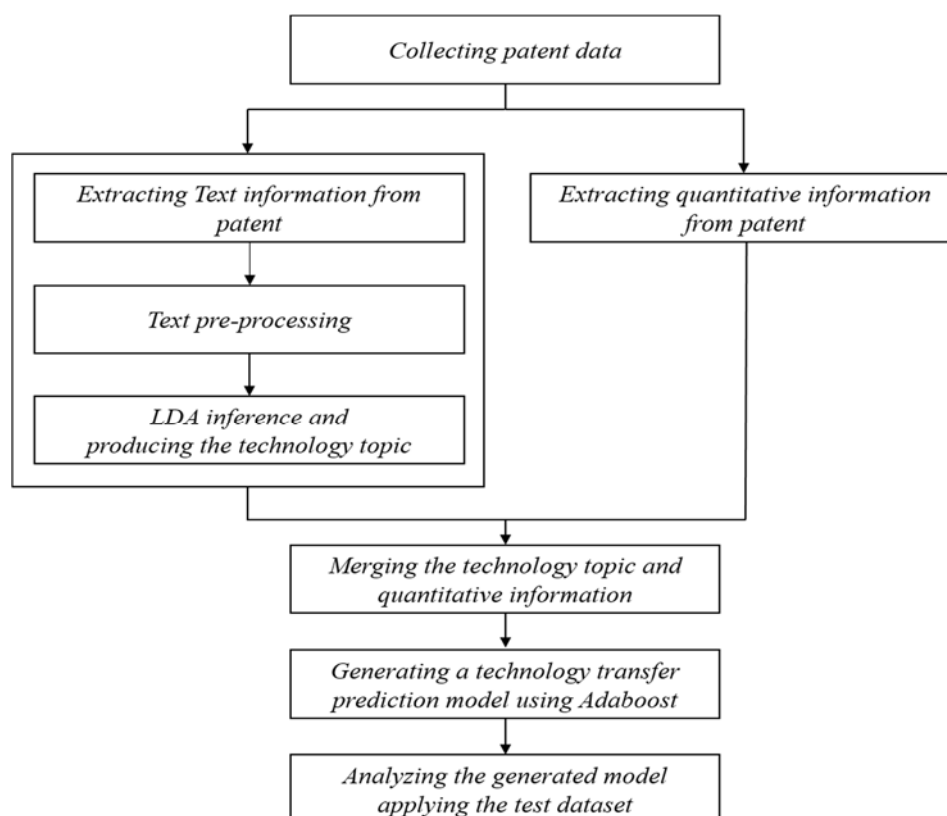


Figure 3. Methodology.

3.1. Collecting the Data

The aim of our study is to produce a technology transfer prediction model using patent data. The patents used for the experiment belong to the domain of inference and machine learning technologies filed before August 2016 in the United States. Because artificial intelligence (AI) technology is a fusion of various technologies, such as speech recognition, natural language processing, machine learning, and so on, it is difficult for a company to conduct research and development on all the sub-technologies in that field. Therefore, a lot of enterprises adopt a technology trading strategy. For this reason, we collect the patent data included in the field of inference and machine learning and try to generate the model. The data were collected from worldwide intellectual property service (WIPS), which is a Korean service provider of patent information. The collected data contains noise or redundant patents, so we removed them. As a result, a total of 711 valid patents were selected for the further analysis. In the patent search database that we used, one is able to check the information of applicant and current assignee of a patent. If current assignee is different from the original applicant, we considered that the patent is transferred. Out of the 711 valid patents, 208 were found to have been transferred. Table 1 shows a detailed description of the collected data.

Table 1. Information on collected data.

Technical Field	Applicant Country	# of Data (# of Transferred)	Period
Inference and machine learning	U.S.	711 (208)	~August 2016

3.2. Technology Clustering Using Topic Model

To enable searchers to locate patents easily, the patent examiner classifies them using IPC, cooperative patent classification (CPC), or file index (FI) according to the contents of the patent

specification. However, since the classification codes do not contain detailed technical content, the ability to ascertain technical material based on them is limited. Therefore, in our study, the technologies in the target domain are first classified through topic modeling.

Figure 4 illustrate the process of finding the technology topic in this research. In order to grasp the technological topics of the collected data, we used the text of patent documents obtained by merging the titles and the abstracts into a corpus, which is a linguistic set of texts. Under the given grammatical structure, the word class is generally written according to the location of the predicate, object, and the like. Among them, stop-words such as “the”, “a”, “by”, “as”, and “is”, referring to commonly used terms, are necessary for building general sentences, but are worthless for the analysis in this study. As such words do not provide special meaning in LDA and cause an unnecessary increase in the complexity of the computation when included, it is necessary to remove them appropriately for efficient information processing. Therefore, preprocessing is performed by eliminating stop-words. In addition, because not every word in the documents is significant for further analysis, the term frequency-inverse document frequency (TF-IDF) method was used to select significant words for the experiment. The form of the word is determined by the position of each word in the sentence. For this reason, if the word form is not the same, a computer is not able to recognize whether some words have the same meaning or not, which can result in a data distortion. Thus, the unification of words that have the same meaning is necessary. In this study, stemming is used to unify the forms of words. Also, we eliminate any numbers, punctuation, and symbols that may distort the analytic process.

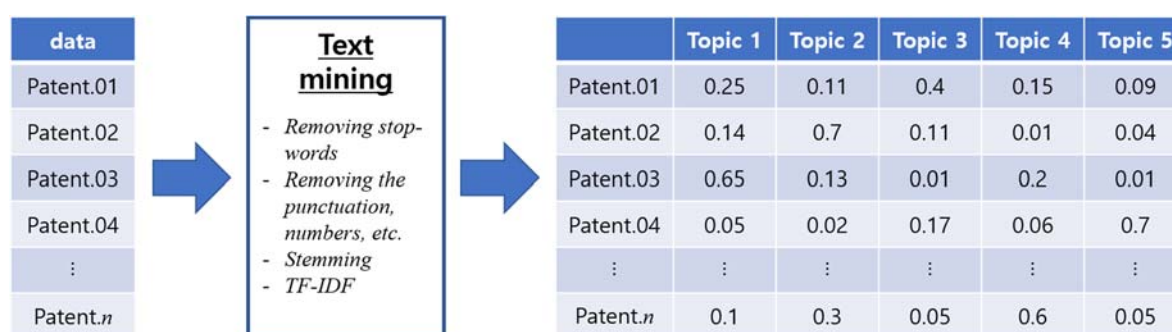


Figure 4. The process of finding the technology topic. TF-IDF—term frequency-inverse document frequency.

The data that have been refined through the preprocessing system are used to fit the LDA model. To fit the LDA model, we adopted the Gibbs sampling method and increased it from 2 to 10 to find the optimal K . In addition, the number of iterations is limited to 500, 1000, and 2000 so that the topic can be displayed well. The Dirichlet parameter α is set as the estimated value from the document and parameter β is set as 0.1. Table 2 shows the most suitable parameters for fitting LDA models found through repetition.

Table 2. Latent Dirichlet allocation (LDA) parameters.

Component	Candidates
Inference Algorithm	Gibbs sampling
The number of K	From 2 to 10
Gibbs sampling iteration	1000
Parameter α, β	$A = T/50, \beta = 0.1$

There are various opinions on how to determine the optimal number of topics [23,24,30–32]. For our purposes, the topic is determined by the topic probability θ of each document.

3.3. Technology Transfer Prediction Model Using the AdaBoost Algorithm

In order to implement the technology transfer prediction model, we use the AdaBoost algorithm, which is a typical ensemble model. Figure 5 illustrates the proposed model, and Table 3 shows the used variables in this study.

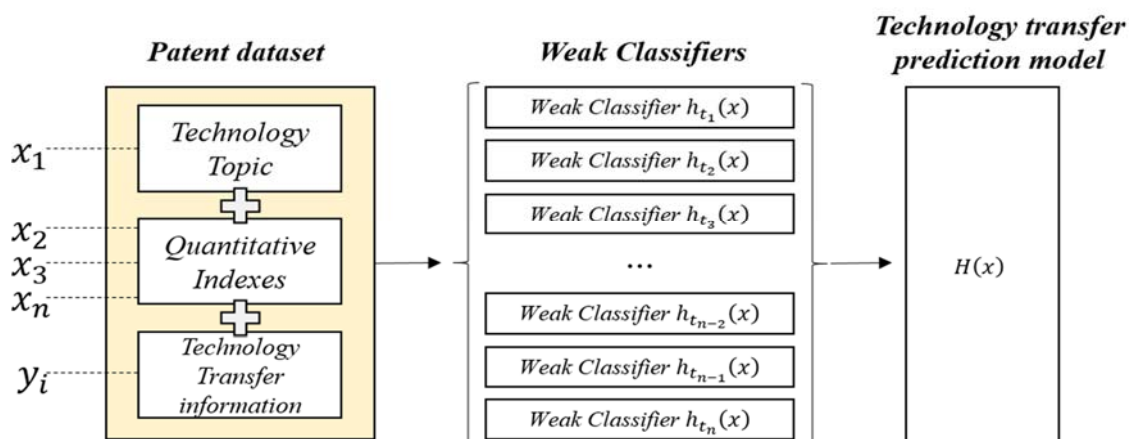


Figure 5. Modeling process for technology transfer prediction model.

Table 3. Variables used in the proposed model.

Variables	Description
Citation	Number of backward citations
Period	The period from application to registration
Claim	Number of registered claims
Family patent	Number of applied family patents
Family country	Number of family countries
Patent references	Number of references (patent)
Non-patent references	Number of references (not patent, e.g., article, book, etc.)
Transfer	Whether a patent has been transferred or not (dummy)
Technology Topic	Topic number according to topic modeling (dummy)

In our experiment, the quantitative indexes mean the variables, such as citation, period, claim, family patent, family country, patent references, and non-patent references. The technology topic represents the technical description of a patent using the LDA. As mentioned in the background, however, it is not appropriate to use categorical independent variables directly as inputs to the AdaBoost algorithm. Therefore, the technology topic variable needs to be changed into the dummy variable. Among the variables in the dataset, “Transfer” indicates whether a patent has been transferred from the original holder to another. In the patent search database that we used, one is able to check the information of applicant and current assignee of a patent. If an assignee is changed, we consider that the patent is transferred. We used “Transfer” as the output variable y_i in this experiment. Therefore, if the patent is transferred to another, it has a value of +1, otherwise it has a value of −1. The parameters used in the AdaBoost algorithm were estimated through iteration. Table 4 shows the parameters used in this experiment.

Table 4. Ensemble model for parameters.

Component	Candidates
Ensemble Model	AdaBoost [27]
Max. depth	4
Iteration	100

To validate the performance of the proposed model, we consider the measures of accuracy, specificity, and sensitivity, which can be used for evaluating the classification performance. The measures are shown in detail in Table 5.

Table 5. Performance measures. TP—true positive; TN—true negative; FP—false positive; FN—false negative; P—positive; N—negative.

Measure	Equation
Accuracy	$\frac{TP+TN}{P+N}$
Sensitivity	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{FP+TN}$

The accuracy represents the overall performance of the classifier; it considers true positive (TP) and true negative (TN) together. However, when a classifier learns the noisy training data excessively, it may cause overfitting. Thus, the accuracy is not able to measure the correct performance of a classifier in such a situation. In order to overcome this problem, we evaluate the performance of the proposed model considering both the sensitivity, which only considers true positive, and the specificity, which considers true negative only.

4. Experiment Result

To carry out this study, we collected data for the technical field of inference and machine learning according to the conditions mentioned at the beginning of Section 3.1.

First, in order to examine the trend of the collected data, we show the trend for applications and technology transfer in the graph in Figure 6. In Figure 6, “n” is the graph showing the number of patent applications by year, and “tr” indicates the number of technology transfers by year.

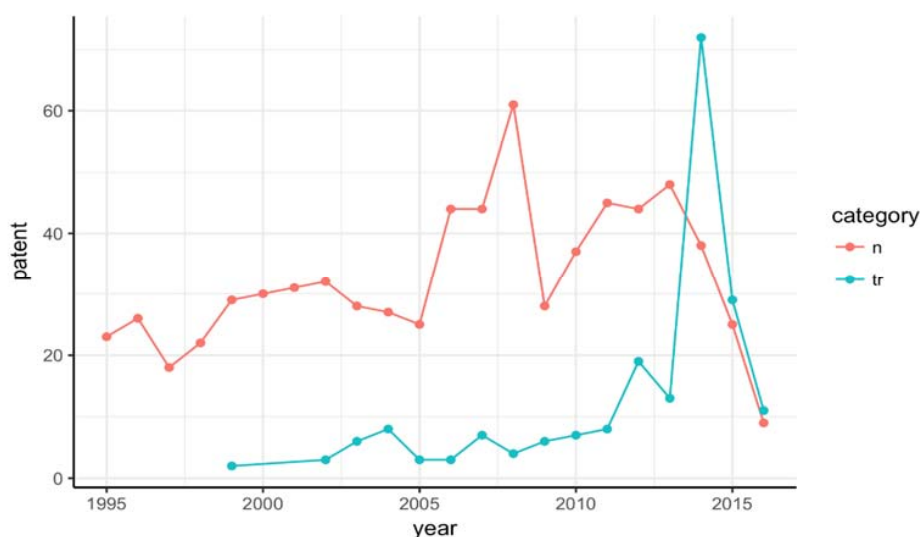


Figure 6. Trend of patent application of part of inference and machine learning and of technology transfer. n—number of patent applications by year; tr—number of technology transfers by year.

As can be seen, from 1995 to 2008, patent applications for the above technology field showed a steady increase. In 2009, patent applications dropped sharply compared with 2008, but the number of applications increased until 2013. However, from 2014 onward, patent applications have decreased again. Looking at the trends in technology transfer, the chart begins to measure it in 1997. Although the number of technology transfers is not large, it has gradually increased since 2008 and rose sharply in 2014.

We would like to know trends according to specific technologies, but the patent data do not have technical classification information. Therefore, in this study, a technical classification is performed using the LDA as mentioned above. The LDA uses the parameters shown in Table 2. In addition, several methods for selecting the optimal number of topics have already been proposed [23,24,31–33]. We used the method Cao et al. proposed, which is considered to be the most appropriate method for this study [31]. The result is shown in Figure 7.

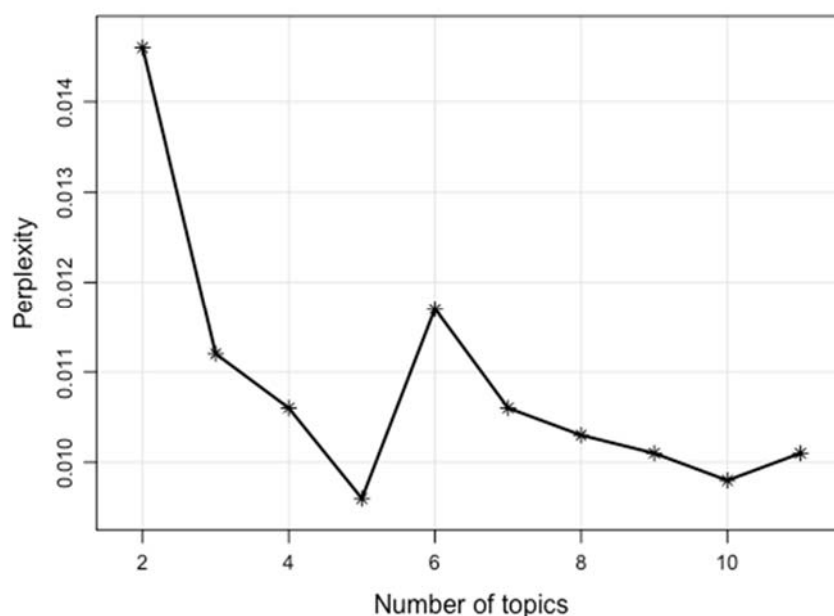


Figure 7. Result of analysis of the number of optimal topics.

As a result of the analysis of the optimal number of topics using the method described by Cao et al., it was found to have a minimum at $k = 5$. Based on these results, we classified the technology topic of collected data. Table 6 shows the results. The top ten keywords included in each topic were utilized to define the technology topic.

Table 6. Technology topics. ANN—artificial neural network.

Topic	Core Keyword	Defined Technology	Quantity
1	node, vector, processor, well, target, patient, expert, complet, element, format	Expert System	152
2	signal, sensor, pattern, circuit, tabl, robot, audio, alarm, power, diagnost	Signal Processing	138
3	reason, unit, sequnc, semant, content, context, languag, score, search, observ	Natural Language Understanding	159
4	imag, object, event, action, manag, service, convolute, filter, agent, fact	Image Processing	144
5	ANN, neuron, modul, layer, extract, fuzzi, document, recur, attribute, simul	Artificial Neural Network	118

As a result of the technical classification, the technology of “natural language understanding” occupied the largest part, about 22.4%, in the field of inference and machine learning. Next, “expert system” technology accounted for 21.4%, followed by signal processing, image processing, and artificial neural network technology. Table 7 shows the number of technology transfers for the above technology fields.

Table 7. Technology transfer according to technology topic.

	Topic 1.	Topic 2.	Topic 3.	Topic 4.	Topic 5.	Total
Not Transferred	112	108	98	100	85	503
Transferred	40	30	61	44	33	208
Ratio of Transferred	26%	22%	38%	31%	28%	29%

The transfer rate of the collected data was 29% on average. As a result of confirming the proportion of technology transfers according to the technology classification using LDA, the rate of technology transfer in the field of Topic 3 (natural language understanding) is 38%, which is relatively higher than other technology fields.

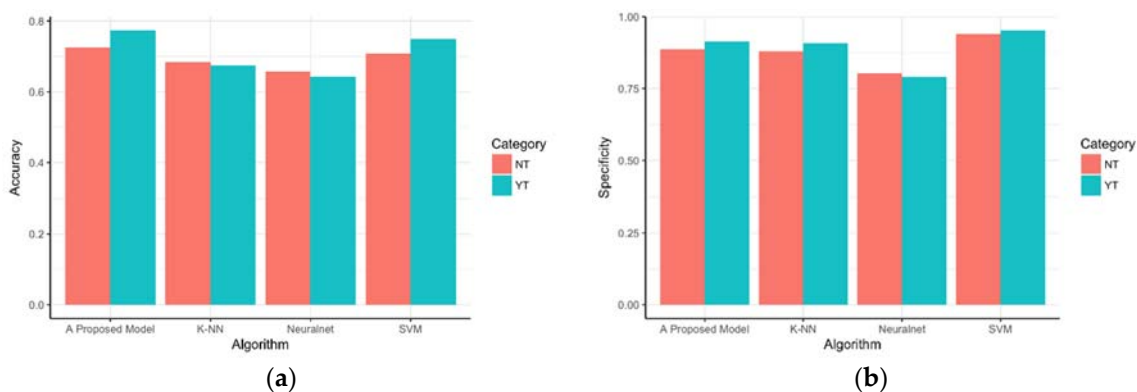
To generate the technology transfer prediction model proposed in this study, we merge the previous topic model results and the quantitative index for patents, which is described in Table 3 and used the AdaBoost algorithm. In order to compare the performance of the proposed method, performance comparison tests are performed with K-nearest neighbor classifier (K-NN), support vector machine (SVM), and neural network algorithm, which are representative classification algorithms. The performance measures use the accuracy, sensitivity, and specificity mentioned in Section 3. We also discuss models that include technology topics and those that do not. The experimental results are shown in Tables 8 and 9 and Figures 8 and 9. In Figures 8 and 9, NT refers to a model that does not include the technology topic, and YT refers to a model that includes the technology topic.

Table 8. Performance measure of proposed model (include the technology topic information). K-NN—K-nearest neighbor classifier, SVM—support vector machine.

	The Proposed Model	K-NN	SVM	Neural Network
Accuracy	0.774	0.675	0.750	0.643
Sensitivity	0.435	0.113	0.258	0.343
Specificity	0.913	0.907	0.953	0.79

Table 9. Performance measure of proposed model (not including the technology topic information).

	The Proposed Model	K-NN	SVM	Neural Network
Accuracy	0.726	0.684	0.708	0.657
Sensitivity	0.339	0.210	0.145	0.357
Specificity	0.887	0.880	0.940	0.804

**Figure 8.** (a) The accuracy of a proposed model and others; (b) The specificity of a proposed model and others. NT—model that does not include the technology topic; YT—model that includes the technology topic. K-NN—K-nearest neighbor classifier, SVM—support vector machine.

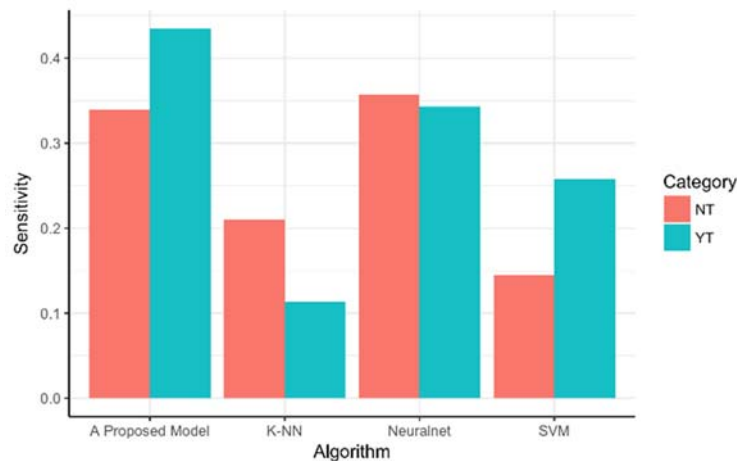


Figure 9. The sensitivity of proposed model and others.

We compared models that include the technology topic with ones that do not include it. The result is shown in the following. The classification performance of the technology transfer prediction model including the technology information from the literature is superior to the model that does not include the literature information. The sensitivity of the model that does not contain the technical content is notably lower than that of the model that includes the technical information because overfitting occurs in the model. Therefore, it can be assumed that technology information is an important factor in predicting technology transfer.

Next, we generated models using the proposed method and other models for comparison, which were based on classifiers such as K-nearest neighbor classifier (K-NN), support vector machine (SVM), and neural network, respectively. The K-NN is simple in structure but has an excellent performance. For this reason, it is used in many classification problems [33–36]. The support vector machine and the neural network are also well known to have excellent classification performance and are applied in various fields [37–40]. As a result of the comparison between the proposed model in this study and the other classifiers mentioned above, the accuracy of the models was found to be similar overall. However, in terms of the sensitivity and the specificity, which indicates the true positive and the true negative, respectively, there was a significant difference between the proposed model and the other models. In particular, the sensitivity and specificity of the other models we compared were lower than those of the proposed model. This seems to be because of the overfitting of the model. These results show that the proposed model performs better than the other models do in the case of technology transfer prediction. It can, therefore, be inferred that the proposed model based on patent data in this study is suitable for predicting technology transfer.

5. Discussion

The advancement of science and technology has made human life more convenient than ever, but competition in society has also become very intense. In today's technology-intensive market environment, companies strive to survive through sustained growth. Such efforts are made in a variety of ways, such as self-driving car alliance. Technology transfer is also a management strategy for maintaining technological competitiveness and sustaining the growth of companies. Previously, studies using surveys or patent data have been conducted to promote technology transfer, but no systematic model was suggested.

This study proposes a predictive model of technology transfer based on an ensemble method to support the continuous growth of enterprises and countries. The proposed model can predict the transferability of patents. In the experimental results, the proposed model showed better classification performance than the other models. If companies or research institutes use the predicted results, it is

possible to select patents with a high potential for being transferred, which can increase the success rates of the transactions. The capital acquired through technology transfer can be reinvested in the activities for continuous growth of enterprises or research institutes.

In future work, we expect to improve the generalization performance of the model by using various technical data. In addition, it is necessary to develop an additional algorithm that is able to enhance the performance of prediction.

6. Conclusions

In recent years, intellectual property has become an indispensable element for the sustainable growth of a corporation. A lot of technology-intensive companies have tried to maintain their competitiveness directly through technological research and development. As a method of technological development is the pursuit of convergence, these days, it is difficult to secure competitiveness using traditional methods. As an alternative, technology transfer is becoming widely used. Technology transfer should be encouraged because not only companies, but also universities and research institutes that have developed technologies, are able to acquire opportunities to create new technologies through such transfers.

Previous work has been studied to find the factors needed to predict technology transfers. In addition, models that only consider either quantitative elements or technical content have been proposed. However, these studies have not focused on a technology transfer prediction model that considers both of them. In this study, we proposed a methodology for predicting technology transfer to enable more effective technology transfers. LDA was used to take into account the technical contents of the collected patent data. Its results were used as variables for the proposed model to represent patent technologies. Also, quantitative factors of the patents were extracted. The technology transfer prediction model based on both the result of the LDA and the quantitative patent variables was finally produced using the AdaBoost algorithm, which is a representative ensemble method. As a result, it was confirmed that the accuracy, sensitivity, and specificity of the proposed model were superior to those of the other methods we compared.

Through the outcome of our study, we were able to predict technology transfer, and the following advantages are expected. There are differences in the quantitative elements of patents in each technology area. Therefore, there is a limitation in that it is challenging to generalize when attempting a technology transfer prediction using only quantitative patent factors. The proposed model is able to reflect the technology field, and it is also sufficient to cover the difference of quantitative factors existing in each technical area as information on technology is included. Also, it is expected that the result of the technology transfer prediction can be taken into consideration when evaluating the value of a technology.

Author Contributions: J.L., J.-H., and S.P. conceived and designed the experiments; D.J., H.L., and S.J. analyzed the data to show the validity of this study; J.L. wrote the paper and performed the entire research step. In addition, all authors have cooperated with each other in revising the paper.

Funding: This research received no external funding.

Acknowledgments: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2015R1D1A1A01059742). This research was also supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of ICT & Future Planning (NRF-2017R1A2B1010208). Lastly, this research was supported by the BK 21 Plus (Big Data in Manufacturing and Logistics Systems, Korea University).

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

References

1. Schilling, M.A. *Strategic Management of Technological Innovation*, 5th ed.; McGraw-Hill Education: New York, NY, USA, 2016; ISBN 9781259539060.
2. Betz, F. *Managing Technological Innovation: Competitive Advantage from Change*, 3rd ed.; Wiley-Interscience: Hoboken, NJ, USA, 2011; ISBN 0470547820.
3. Rivette, K.G.; Nothhaft, H.R.; Kline, D. Discovering New Value in Intellectual Property, 2000. Available online: <https://hbr.org/2000/01/discovering-new-value-in-intellectual-property> (accessed on 11 April 2018).
4. Roper, A.T.; Cunningham, S.W.; Porter, A.L.; Mason, T.W.; Rossini, F.A.; Banks, J. *Forecasting and Management of Technology*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2011; ISBN 9780470440902.
5. Abraham, B.P.; Moitra, S.D. Innovation assessment through patent analysis. *Technovation* **2001**, *21*, 245–252. [[CrossRef](#)]
6. Jun, S.; Park, S.S. Examining technological innovation of Apple using patent analysis. *Ind. Manag. Data Syst.* **2013**, *113*, 890–907. [[CrossRef](#)]
7. Daim, T.U.; Rueda, G.; Martin, H.; Gerdtsri, P. Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technol. Forecast. Soc. Chang* **2006**, *73*, 981–1012. [[CrossRef](#)]
8. Jun, S.; Park, S. Examining technological competition between BMW and Hyundai in the Korean car market. *Technol. Anal. Strateg. Manag.* **2016**, *28*, 156–175. [[CrossRef](#)]
9. Kim, G.; Bae, J. A novel approach to forecast promising technology through patent analysis. *Technol. Forecast. Soc. Chang* **2017**, *117*, 228–237. [[CrossRef](#)]
10. Mello, J.P. Technology Licensing and Patent Trolls. *BUJ Sci. Tech. L* **2006**, *12*, 388.
11. Lee, J.; Um, C. Innovation strategy of mobile industry in Korea: Case study of CDMA. In Proceedings of the 2006 IEEE Portland International Conference on Management of Engineering and Technology, Istanbul, Turkey, 8–13 July 2006; Volume 5, pp. 2052–2060.
12. Bozeman, B. Technology transfer and public policy: A review of research and theory. *Res. Policy* **2000**, *29*, 627–655. [[CrossRef](#)]
13. Choi, J.; Jang, D.; Jun, S.; Park, S. A Predictive Model of Technology Transfer Using Patent Analysis. *Sustainability* **2015**, *7*, 16175–16195. [[CrossRef](#)]
14. Trappey, A.J.C.; Trappey, C.V.; Wu, C.-Y.; Lin, C.-W. A patent quality analysis for innovative technology and product development. *Adv. Eng. Inform.* **2012**, *26*, 26–34. [[CrossRef](#)]
15. Park, S.; Jun, S. Technology Analysis of Global Smart Light Emitting Diode (LED) Development Using Patent Data. *Sustainability* **2017**, *9*, 1363. [[CrossRef](#)]
16. Lai, W.H.; Tsai, C.T. Fuzzy rule-based analysis of firm's technology transfer in Taiwan's machinery industry. *Expert Syst. Appl.* **2009**, *36*, 12012–12022. [[CrossRef](#)]
17. Park, Y.; Lee, S.; Lee, S. Patent analysis for promoting technology transfer in multi-technology industries: The Korean aerospace industry case. *J. Technol. Transf.* **2012**, *37*, 355–374. [[CrossRef](#)]
18. Park, H.; Yoon, J.; Kim, K. Expert Systems with Applications Using function-based patent analysis to identify potential application areas of technology for technology transfer. *Expert Syst. Appl.* **2013**, *40*, 5260–5265. [[CrossRef](#)]
19. Yoon, B.; Park, Y. A text-mining-based patent network: Analytical tool for high-technology trend. *J. High Technol. Manag. Res.* **2004**, *15*, 37–50. [[CrossRef](#)]
20. Yoon, B.; Park, Y. A systematic approach for identifying technology opportunities: Keyword-based morphology analysis. *Technol. Forecast. Soc. Chang* **2005**, *72*, 145–160. [[CrossRef](#)]
21. Kyebambe, M.N.; Cheng, G.; Huang, Y.; He, C.; Zhang, Z. Forecasting emerging technologies: A supervised learning approach through patent analysis. *Technol. Forecast. Soc. Chang* **2017**, *125*, 236–244. [[CrossRef](#)]
22. Kim, J.M.; Jun, S. Zero-inflated poisson and negative binomial regressions for technology analysis. *Int. J. Softw. Eng. Appl.* **2016**, *10*, 431–448. [[CrossRef](#)]
23. Grün, B.; Hornik, K. topicmodels: An R Package for Fitting Topic Models. *J. Stat. Softw.* **2011**, *40*, 1–30. [[CrossRef](#)]
24. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
25. Kim, G.; Park, S.; Jang, D. Technology Analysis from Patent Data Using Latent Dirichlet Allocation. In *Soft Computing in Big Data Processing*; Lee, K.M., Park, S.-J., Lee, J.-H., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 71–80.

26. Lee, H.; Kang, P. Identifying core topics in technology and innovation management studies: A topic model approach. *J. Technol. Transf.* **2017**. [[CrossRef](#)]
27. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1995**, *139*, 23–37. [[CrossRef](#)]
28. Alfaro, E.; Gamez, M.; García, N. Adabag: An R Package for Classification with Boosting and Bagging. *J. Stat. Softw.* **2013**, *54*, 1–35. [[CrossRef](#)]
29. Schapire, R.E.; Singer, Y. Improved Boosting Algorithms Using Confidence-rated Predictions. *Computer* **1999**, *336*, 297–336. [[CrossRef](#)]
30. Griffiths, T.L.; Steyvers, M. Finding scientific topics. *Proc. Nat. Acad. Sci. USA* **2004**, *101*, 5228–5235. [[CrossRef](#)] [[PubMed](#)]
31. Cao, J.; Xia, T.; Li, J.; Zhang, Y.; Tang, S. A density-based method for adaptive LDA model selection. *Neurocomputing* **2009**, *72*, 1775–1781. [[CrossRef](#)]
32. Arun, R.; Suresh, V.; Madhavan, C.V.; Murthy, M.N. On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations. In *Advances in Knowledge Discovery and Data Mining*; Zaki, M.J., Yu, J.X., Ravindran, B., Pudi, V., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 391–402.
33. Kwartler, T. *Text Mining in Practice with R*, 1st ed.; Wiley: Hoboken, NJ, USA, 2017; ISBN 9781119282013.
34. Han, E.-H.S.; Karypis, G.; Kumar, V. Text Categorization Using Weight Adjusted k-Nearest Neighbor Classification. In *Advances in Knowledge Discovery and Data Mining*; Cheung, D., Williams, G.J., Li, Q., Eds.; Springer: Berlin/Heidelberg, Germany, 2001; pp. 53–65.
35. Bang, S.L.; Yang, J.D.; Yang, H.J. Hierarchical document categorization with k-NN and concept-based thesauri. *Inf. Process. Manag.* **2006**, *42*, 387–406. [[CrossRef](#)]
36. Versus, S.; Methods, U. k-nearest neighbor algorithm. In *Discovering Knowledge in Data: An Introduction to Data Mining*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2005; pp. 90–106. ISBN 0471666572.
37. Lam, S.L.Y.; Lee, D.L. Feature reduction for neural network based text categorization. In Proceedings of the 6th International Conference on Database Systems for Advanced Applications, Hsinchu, Taiwan, 19–21 April 1999; pp. 195–202. [[CrossRef](#)]
38. Yu, B.; Xu, Z.B.; Li, C.H. Latent semantic analysis for text categorization using neural network. *Knowl. Based Syst.* **2008**, *21*, 900–904. [[CrossRef](#)]
39. Joachims, T. Text categorization with Support Vector Machines: Learning with many relevant features. In *Machine Learning: ECML-98*; Nédellec, C., Rouveirol, C., Eds.; Springer: Berlin/Heidelberg, Germany, 1998; pp. 137–142.
40. Drucker, H.; Wu, D.; Vapnik, V.N. Support vector machines for spam categorization. *IEEE Trans. Neural Netw.* **1999**, *10*, 1048–1054. [[CrossRef](#)] [[PubMed](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).