PAPER

Patent Registration Prediction Methodology Using Multivariate Statistics

Won-Gyo JUNG[†a)], *Member*, Sang-Sung PARK[†b)], *and* Dong-Sik JANG[†c)], *Nonmembers*

SUMMARY Whether a patent is registered or not is usually based on the subjective judgment of the patent examiners. However, the patent examiners may determine whether the patent is registered or not according to their personal knowledge, backgrounds etc. In this paper, we propose a novel patent registration method based on patent data. The method estimates whether a patent is registered or not by utilizing the objective past history of patent data instead of existing methods of subjective judgments. The proposed method constructs an estimation model by applying multivariate statistics algorithm. In the prediction model, the application date, activity index, IPC code and similarity of registration refusal are set to the input values, and patent registration and rejection are set to the output values. We believe that our method will contribute to improved reliability of patent registration in that it achieves highly reliable estimation results through the past history of patent data, contrary to most previous methods of subjective judgments by patent agents.

key words: *patent, neural network, pattern recognition, data mining, text mining*

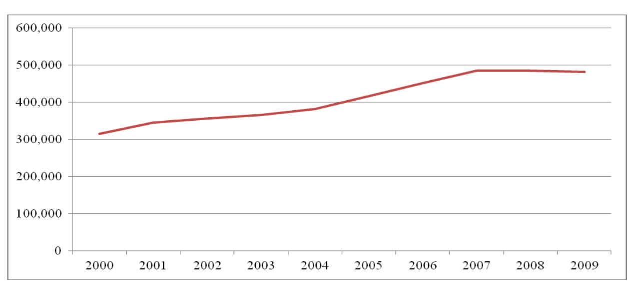
1. Introduction

A patent is a set of exclusive rights granted to an inventor or their assignee for a limited period of time in exchange for public disclosure of an invention [1]. The patent grant is usually reviewed by the patent examiners after its specification is applied [2]. However, for granting the patent, a patent application should fulfill the specific conditions prescribed in patent law. The patent application must include one or more claims defining the invention based on novelty, inventive steps, and industrial application [3]. Inventions prescribed in patent law mean not only technology, but also advanced technological ideas based on laws of nature. A technological idea is an abstract and conceptual idea which does not reach a certain level of practical use, whereas a technology is a specific tool utilized in real industry. Whether inventions are based on laws of nature or not is determined by the claims written in the patent applications. Inventors do not have to prove the laws of nature. They can simply use the laws of nature based on their experience. Also, inventions must show a certain level of creativity [4].

A "novelty" means that the claims defining the invention should not be publically released. The principle that a patent must be new was prescribed in England (1623) and has become one of the basic principles for the patent system [5]. However, even a patent granted to a low-level technology can hinder industrial development, so the patent office prescribes inventive steps for patentability to rule out inventions which do not reach a certain level of practical use as well as to protect creative inventions. Inventions are considered "inventive steps" according to whether they are simply made by precedent technology. Also, since patent law aims to develop industry, a patent which is not applicable for industry is worthless. An "industrial application" means that inventions must be utilized in real industry or will be in the near future. In this paper, we choose input variables which meet the requirements of the patent, such as novelty, inventive steps and industrial application [6].

In particular, patents have become important business interests, so a number of companies invest their resources in patent strategies [7]. Countries have also begun to reinforce patent development policies at the national level, because a patent has become an indicator of technical competitive strength [8]. As the importance of patents increases, the amount of patent applications has also increased. Figure 1 shows the growth rate of patent applications from 2000 to 2009. However, the patent registration ratio is below 50% in spite of the high rate of patent applications, since a lot of human and economic resources are needed to register a patent [9]. Therefore, most companies will operate at a loss when their patent applications are rejected due to similar technologies. To prevent such economic losses, in general, companies research all similar patents beforehand using a keyword search in databases. And this type of patent research is characterized by the subjective judgment of the patent examiners [10].

Therefore, we present a patent registration prediction method to avoid subjective judgments and automate the in-



Source : http://www.uspto.gov

Fig. 1 Increases in patent applications.

Manuscript received July 26, 2010.
Manuscript revised March 23, 2011.

[†]The authors are with the Division of Information Management Engineering, Korea University, Seoul, Korea.
a) E-mail: themong@korea.ac.kr
b) E-mail: hanyul@korea.ac.kr
c) E-mail: jang@korea.ac.kr

DOI: 10.1587/transinf.E94.D.2219

creased patent review work. The proposed patent registration prediction method will estimate whether patent applications are registered or not by implementing a pattern recognition method such as a neural network algorithm. In our prediction method, the application date, activity index, IPC code and similarity of registration refusal are set to the input values, and the patent registration and rejection are set to the output values. Similarity of registration refusal among input values indicates similarities between patent applications and rejected applications. We also experimented with patents in a variety fields, such as Bluetooth, solar-energy, and hard-disk technologies. The reminder of the paper is organized as follows. In Sect. 2, we summarize the related work and survey recent research trends. In Sect. 3, we present a patent registration estimation method based on pattern recognition. In Sect. 4, we empirically test the proposed method and summarize the results. Finally, Sect. 5 concludes this study with the expected effects by proposing directions for future study.

2. Related Literature

Patent data not only indicates the technological contents of the patent itself but also estimates the trends and emergence of new technologies by analyzing the words, citations of previously applied patents and the scope of claims in its patent. From the study of Kutznets (1962), the patent information is considered more useful than other indexes, such as the possibility of gaining data, diversity in the range of data, and amount of information of each technology for economic analysis [11]. Especially, the citation analysis can draw following indexes; technology impact factor, patent life cycle (Hirschey. 2001), economical value in the innovation activities, and technology combination and knowledge transfer between nations (Tijssen. 2001) etc. And it was presented that those indexes can measure quality level of technology asset [12]–[14]. Yoon and Park (2004) extracted topic words which explain the characteristics of patents through data mining, and they found the relation between patents based on the topic words and experimented with replacing the functions of existing patent citations. In this experiment, links are topic words extracted from patents, and duplicated topic words and link distances are processed with a patent index conversion technique. The results confirm that a general patent index can judge the influence, lifespan, and technology duplication of patents as well as CII [15]. Lee (2003) also analyzes the quality and quantity of Korean patents by using patent indexes such as CII, and identifies technological innovation in Korea from 1980 to 2001. This analysis is based on the frequency of the patent index [16]. Yoo (2004) use forward citation classification for American patents to estimate how long a specific technology will continue to have an influence [17]. Tseng (2007) draw patent maps by using data mining techniques, and analyze the relation of each technology. This paper aims to automate the whole process, which not only helps create a final patent map for topic analyses, but also facilitates or improves other patent analysis tasks such as patent classification, organization, knowledge sharing, and prior art searches [18]. Lanjouw and Schankerman (2004) studies a correlation between the patent quality and research productivity from the survey of approximately 100,000 patents over the 7 technology areas, applied by American manufacturing company from 1974 to 1993. As a result of this, it was revealed the patent quality exerts a positive influence upon stock's market value, and the number of claims is the most powerful factor in the area of 6 technologies [19]. Lin, Chen, and Wu (2007) also surveys American patents over the 14 fields of biotechnology and presents that the nationality of applicant, geographical position, the number of claims are statistically associated with cited times. On the other hand, there is a positive or negative correlation between evaluation period and the cited times [20]. Our research differs from existing studies as follows. Firstly, it is the first attempt to estimate the possibility of patent registration based on pattern reorganization. Secondly, at the point of the estimation model, we experimented by comparing our model with other widely used models. From the experiment, we proved that our model's performance is higher than that of the existing models.

3. Methodology

3.1 Neural Network Algorithm

The Neural Network (NN) used the Back Propagation Neural Network (BPN) which is an efficient learning method of the Multi-Layer Perceptron (MLP). The BPN consists of a 3-layered feed-forward, which has a hidden layer between the input and output layers [21].

The BPN learning method involves a learning process that changes the initial connection weight value to a suitable value. The forward stage presents the input pattern of the neural network, and calculates the output using the input and activation functions [22]. The activation function used in the calculation of the connection weight was based on a sigmoid function such as expression (1).

$$\log sig(x) = \frac{1}{1 + \exp(-x)} \tag{1}$$

The output value of the hidden and output layers can by represented by expression (2) and (3).

$$h = \log sig\left(\sum_{i=1}^{n} w_{ij} x_i\right) \tag{2}$$

$$y = \log sig\left(\sum_{j=1}^{m} w_{jk} x_j\right) \tag{3}$$

x_i = input variables
w_{jk} = connection weight between the hidden and output layers
w_{ij} = connection weight between input and hidden layers

The backward stage renews the connection weight which is an important element of learning. And, the error

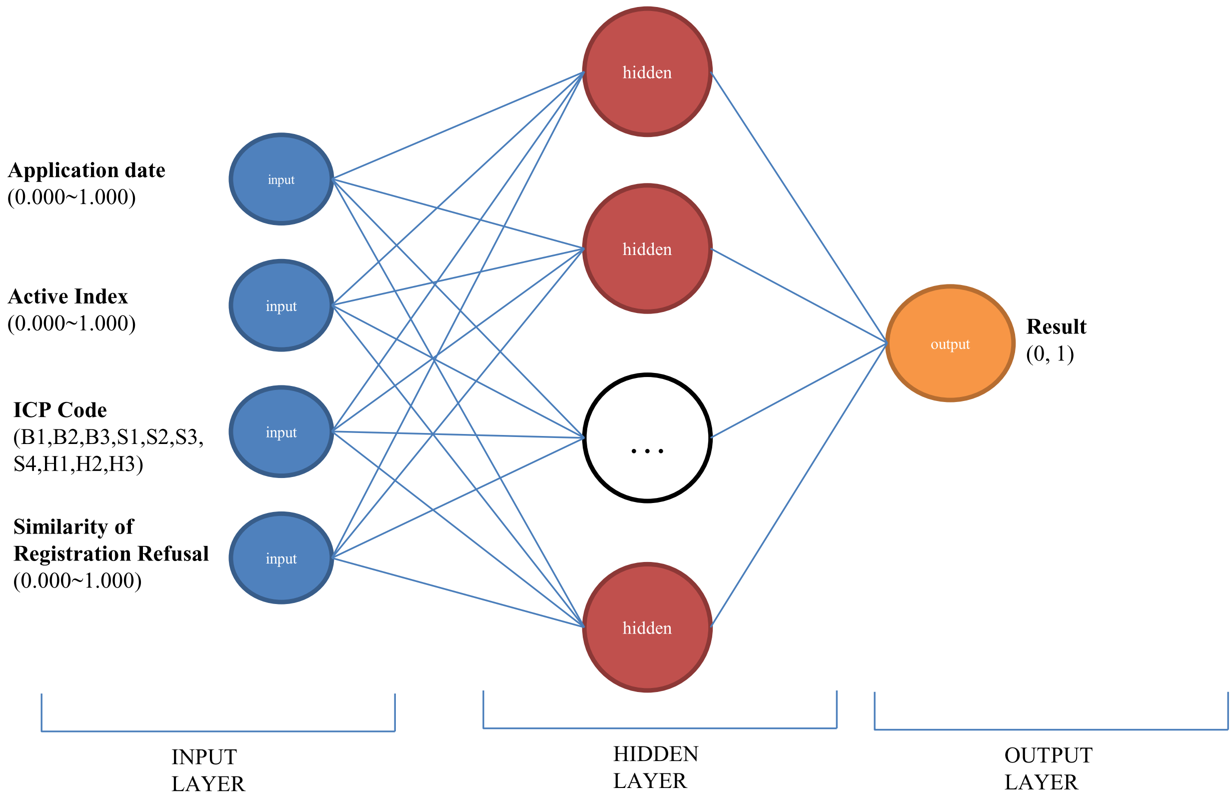


Fig. 2 Graphical representation of our NN model.

between the target and output values is found by calculating expression (4), and then the connection weight between layers is sequentially renewed from the output to input layers in order to minimize the error value. Through these stages, it finds the final connection weight which minimizes the error.

$$e = \frac{\sum_{n=1}^{j}(y-t)^2}{2} \tag{4}$$

We designed the number of input layers, hidden layers and output layers in the NN as 4, 12, and 1 in each representation. The graphical representation of our NN is shown in Fig. 2.

3.2 Performance Metrics

To estimate the success ratio of patent registration, the PSR (Percent Success Rate) is chosen. The PSR is an intuitive statistical method which identifies the estimation accuracy. And, the PSR represents the number of samples correctly identified divided by the total number of samples. The bigger the PSR, the better the performance. The PSR can be formulated as in Eq. (5).

$$PSR = \frac{\textit{Number of samples correctly identified}}{\textit{Total number of samples}} \tag{5}$$

Table 1 Summary of independent variables.

Independent variable name	Number of values	Possible values
Application Date Index	Range	0.000~1.000
Activity Index	Range	0.000~1.000
Division of IPC Code	10	B1, B2, B3, S1, S2, S3, S4, H1, H2, H3
Similarity of Registration Refusal	Range	0.000~1.000

3.3 Data and Variable Definitions

To verify the performance of our model, 200 registered patents and 200 rejected patent documents are chosen for each Bluetooth, solar cell, and hard disk technology. And, the total amount of data, used to construct the database for our system was 12,000 documents. In the estimation model, the patent date, IPC code, and similarity of registration refusal are set as independent variables, and the registration and rejection are set as dependent variables. For the study of each technology, 350 training and 50 testing documents are used. The summary of each independent variable is shown in Table 1 along with a specific explanation.

3.3.1 Application Date Index

The application date refers to the date that the patents are applied. Regarding the patents, the applied date represents the time when the technologies are invented. Among the patent requirements, “new” is determined based on the registered date. The registered date is an important determining factor, because patents can be rejected if similar technologies are found before the registered date. In order to apply to our estimation model, we generated random values of registered data, called Application Date Index (ADI), ranging from 0 to 1. ADI can be formulated as in Eq. (6).

$$ADI = \frac{Application\ date - Min(Application\ date)}{Max(Application\ date) - Min(Application\ date)} \tag{6}$$

3.3.2 Activity Index

Activity Index (AI) represents an index to identify the relative impact of the patent on the applicant’s specific technology areas. If the AI is bigger than 1, it means relative patent activity is vigorous. The AI can be formulated as in Eq. (7). In order to apply to our estimation model, we generated random values of AI ranging from 0 to 1.

$$AI = \frac{\dfrac{The\ number\ of\ applicant's\ specific\ technical\ areas}{Total\ number\ of\ specific\ technical\ areas}}{\dfrac{Total\ number\ of\ applicants}{Total\ number\ of\ patents}} \tag{7}$$

3.3.3 Division of IPC Code

The IPC Code (International Patent Classification Code) is a patent classification system for use by any country. The IPC code is presently using the eighth edition and is undergoing continuous revision, The first edition was published in 1968. The IPC Code undergoes revision according to technology development because new technology patents are being continuously applied. Each ‘Section’ of the IPC Code is classified in ‘Table 2’, and reported by stage of ‘section→class →subclass→maingroup→subgroup’.

The IPC Code generally uses variables for patent technology classification. Based on the data used in this paper, the IPC code for Bluetooth technology is divided into three groups we named B1, B2 and B3. The IPC code for solar battery technology is divided into four groups we named S1, S2, S3 and S4. The IPC code for hard disk technology is divided into three groups we named H1, H2 and H3. The IPC code is classified according to Table 3.

Table 2 Section of IPC code (WIPO).

Section	Contents
A	Human necessities
B	Performing operations; Transporting
C	Chemistry; Metallurgy
D	Textiles; Paper
E	Fixed Constructions
F	Mechanical engineering; Lighting; Heating; Weapons; Blasting
G	Physics
H	Electricity

3.3.4 Similarity of Registration Refusal

Similarity of registration refusal is a variable that represents the level of similarity based on patent application refusals. A text mining technique is applied to determine the similarity of registration refusal [23]. Each word written in Abstract, Title, Description, Claims of refused patents is arrayed according to its frequency of use, and unnecessary words, e.g. ‘a’, ‘the’, ‘and’, ‘but’, ‘in’, ‘at’, etc., are eliminated in order to calculate the frequency of each core word [24]. Figure 3 presents the extraction algorithm which calculates frequency of the core words.

After the frequency of the core words are calculated, the weights of the core words are calculated. The generation algorithm for the weights of the core words is shown in Fig. 4.

The similarity value summarizes the results by multiplying by P^*_{iy} based on the results of the Extraction Algorithm for Core Words from unused patent documents in order to construct an integrated DB of core words by using V_i which is the weight of core words in the integrated DB.

Table 3 Division of IPC code.

Technology	Division of IPC Code	IPC Code	Contents
Bluetooth	B1	H04B	Transmission
	B2	H04L	Transmission of digital information, e.g. telegraphic communication
	B3	G06F	Electric digital data processing
Solar	S1	H01L	Semiconductor devices; Electric solid state devices not otherwise provided for
	S2	H02N	Electric machines not otherwise provided for
	S3	C25B	Electrolytic or electrophoretic processes for the production of compounds or non-metals; Apparatus therefor
	S4	F03G	Spring, weight, inertia, or like motors; Mechanical-power-producing devices or mechanisms, not otherwise provided for or using energy sources not otherwise provided for
Hard Disk	H1	G06F	Electric digital data processing
	H2	G11B	Information storage based on relative movement between record carrier and transducer
	H3	H05K	Printed circuits; Casings or constructional details of electric apparatus; Manufacture of assemblages of electrical components

y = Set of patent document

i = Set of words in document y

Step 1. $P_{iy} = \dfrac{W_{iy}}{\sum_{i=1}^{n} W_{iy}}$ (i=1, 2, 3, ... n)

P_{iy} = Word frequency rate of i in document y

W_{iy} = Word frequency of i in document y

Step 2. $C_y = \dfrac{1}{n}$

n = the number of whole words in document y

C_y = Extracting criterion value of core word in document y

Step 3. Eliminate whole word which is $P_{iy} < C_y$

P_{iy} value of rest word are declared to P^{*}_{iy}

Fig. 3 Extraction algorithm for core words.

y = Set of rejected patent document

i = Set of words in document y

Step 1. P^{*}_{iy} = The result of Extraction Algorithm for Core Word

Step 2. $V_i = \dfrac{\sum_{y=1}^{m} P^{*}_{iy}}{R_i}$ (y=1, 2, , ..., m)

V_i = Weight of word i

R_i = The number of duplicated core word

Step 3. If there is no duplicated core word

$V_i = P^{*}_{iy}$

Fig. 4 Generation algorithm for weight of core words.

Figure 5 shows the selection algorithm for similarity value.

The data of application date index, activity index, division of IPC code, and similarity of registration refusal in the Bluetooth technology field is represented in Table 4.

y = Unused patent documents of constructing integration core word DB

i = Duplicated word of integration core word DB and y

Step 1. Get each P^{*}_{iy} value by implementing Extraction Algorithm for Core Word from unused patent documents to construct integration core word DB

Step 2. Select the P^{*}_{iy} value of duplicated word and Weight value V_i by comparing with the integration core word DB

Step 3. $S_y = \sum_{i=1}^{n} (V_i \times P^{*}_{iy})$

V_i = Weight of word i

S_y = Similarity value of document y

Fig. 5 Selection algorithm for similarity value.

Table 4 Value of variables in the bluetooth technology fields.

Application Date Index	Activity Index	Division of IPC Code	Similarity of Registration Refusal	Registration
0.714	0.431	B1	0.052	0
0.429	0.842	B2	0.097	0
0.857	0.423	B1	0.395	1
0.571	0.124	B3	0.053	0
0.571	0.543	B2	0.047	0
0.429	0.679	B2	0.128	1
0.714	0.438	B2	0.138	1
0.286	0.187	B3	0.112	1
0.571	0.135	B1	0.015	0
0.429	0.157	B3	0.084	0
0.429	0.984	B1	0.094	1
0.429	0.842	B1	0.253	1
0.571	0.541	B2	0.116	1
…	….	…	…	…

4. Experiments and Results

4.1 NN Performance

The experiment results of our algorithm are shown in Table 5. For the experiments, 50 testing documents including registered patent applications and patent application refusals were selected. Firstly, in the Bluetooth field, 27 registered patent applications were estimated to be registered and 4 registered patent applications were estimated to be refused. And, 16 patent application refusals were estimated to be refused and 3 patent application refusals were estimated to be registered. Secondly, in the solar cell field, 21 registered patent applications were estimated to be registered

Table 5 Confusion matrix for NN prediction results.

Technology	Bluetooth		Solar cell		Hard disk	
Actual / Predict	Reg.	Ref.	Reg.	Ref.	Reg.	Ref.
Reg.	27	4	21	3	17	2
Ref.	3	16	1	25	4	27

(Reg. : Registration, Ref. : Refusal)

Table 6 PSR for NN prediction results.

Bluetooth	Solar cell	Hard disk
0.86	0.92	0.88

and 3 registered patent applications were estimated to be refused. And, 25 patent application refusals were estimated to be refused and 1 patent application refusal was estimated to be registered. Finally, in the hard disk field, 17 registered patent applications were estimated to be registered and 2 registered patent applications were estimated to be refused. And, 27 patent application refusals were estimated to be refused and 4 refusal patent applications were estimated to be registered.

The PSR was calculated to verify the performance of our model. The PSR was 0.86, 0.92, and 0.88 for the respective Bluetooth, solar cell, and hard disk technology, and the average for the three technologies was 0.89. The PSR for the NN prediction results are summarized in Table 6.

4.2 Comparison to Other Models

We compared our model with other classification methods using the same data. Specifically, we used traditional statistical classification methods, such as the k-means algorithm and logistic regression, and a discriminant analysis. Short descriptions for each of these three classification methods follow.

K-means is a well known simple unsupervised learning algorithm. K-means calculates the distance of input data using criteria based on a preset number of clusters and a minimum cluster size. The algorithm is generally used because of its fast convergence. Normally, far less repetition is needed compared to the initial number of data. In terms of efficiency it does not guarantee an optimal value in that the wrong result can be obtained by changing the initial value. However, the right result can be obtained when another initial value is applied. The only weakness of this algorithm is that the k value needs to be set, and if the distribution of the data is not natural, the wrong result can be obtained [25].

Logistic regression is a generalization of linear regression. It is used primarily for predicting binary or multi-class dependent variables. Because the response variable is discrete, it cannot be modeled directly by linear regression. Therefore, rather than predicting the point estimate of the event itself, it builds a model to predict the odds of its occurrence. In a two-class problem, odds greater than 50% refer to cases assigned to class ‘1’ or class ‘0’ otherwise. While logistic regression is a very powerful modeling tool,

Table 7 Performance comparison of forecasting techniques.

Folds	k-means	Logistic Regression	Discriminant Analysis	Neural Network
1	0.84	0.83	0.74	0.89
2	0.86	0.84	0.75	0.91
3	0.88	0.81	0.77	0.92
4	0.85	0.85	0.81	0.88
5	0.81	0.79	0.76	0.88
6	0.78	0.81	0.75	0.89
7	0.83	0.84	0.71	0.92
8	0.85	0.80	0.82	0.87
9	0.87	0.85	0.76	0.86
10	0.83	0.81	0.77	0.85
Mean	0.84	0.82	0.76	0.89
StDev	0.03	0.02	0.03	0.02

it assumes that the response variable is linear in the coefficients of the predictor variables. Furthermore, the modeler, based on his or her experience with the data and data analysis, must choose the right inputs and specify their functional relationship to the response variable [26].

Discriminant analysis is one of the oldest statistical classification techniques, first introduced by Fisher (1936) [27]. Using the historic data, it finds hyperplanes (e.g. lines in two dimensions, planes in three dimensions, etc.) that separate the classes. The resultant model is very easy to interpret, because it simply determines on which side of the line (or hyperplane) a point falls. Training is simple and scalable. Despite its scalability and simplicity, discriminant analysis is not a popular technique in data mining for two main reasons: (1) it assumes that all of the predictor variables are normally distributed (i.e. their histograms match bell-shaped curves), which may not be the case, and (2) the boundaries that separate the classes are all linear forms (such as lines or planes). However, sometimes the data simply cannot be separated in that manner.

In the following section, we present the results of our method compared to the three classification methods described above. We used exactly the same training and testing data set generated using stratified 10-fold cross validation for all three models and the NN model. The aggregated results are shown in Table 7. As the results indicate, on average, the NN model generated significantly better classification accuracy than all the other methods.

5. Discussion and Conclusion

The experiment results show that the accuracy of our model to estimate patent registration is approximately 89%. Under the same experiment environments, the accuracy of our method is higher compared to k-means, logistic regression and discriminant analysis. Our method is used to estimate whether applied patents are registered or not before they are judged.

Our method will assist companies willing to estimate whether their applied patents are registered or not. Our study is also valuable in that it proved that neural pattern reorganization is worthwhile for problems which are difficult to estimate.

Other methods of patent analysis such as Yoon (2004), Lee (2003) and Tseng (2007) only construct patent maps and analyze the relation between patents. However, our method enables early estimation in order to construct a strategy of patent registration. We found that a hybrid method provides better information for understanding technical contents and constructing technical strategies.

A neural network algorithm is also constructed from arbitrary sets of probabilistic modeling methods. The algorithm constructs vector maps and then adjusts the weights according to its parameters, such as the rate of study, hidden layers etc. The right selection of parameters plays an essential role in the development of neural network models.

To select the right parameters, the model designer's experience and intuition as well as trial-and-error is needed. Recently, researchers have developed a hybrid architecture applied in gene algorithms and other methods of intelligent search in order to optimize the parameters. It is reported that the hybrid architecture greatly influences the selection of parameters. This hybrid architecture can be considered for future work, and, to improve the reliability of our testing, more patent data needs to be included in the experiments.

Acknowledgments

This work was supported by the Brain Korea 21 Project in 2011.

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (MEST) (NRF-R1A4A007-2011-0026953).

References

[1] D. Archibugi and M. Pianta, "Measuring technological change through patents and innovation survey," Technovation, vol.16, no.9, pp.451–468, 1996.

[2] C. Bédécarrax and C. Huot, "A new methodology for systematic exploitation of technology databases," Information Processing & Management, vol.30, no.3, pp.407–418, 1994.

[3] Korea Intellectual Property Office, "Patent and Information Analysis," pp.46–58, 2007.

[4] K. Henning and T. Ulrike, "Chinese regional innovation systems in times of crisis: The case of Guangdong," Asian Journal of Technology Innovation, vol.17, no.2, pp.101–128, 2009.

[5] V.K. Gupta and N.B. Pangannaya, "Carbon nanotubes: Bibliometric analysis of patents," World Patent Information, vol.22, pp.185–189, 2000.

[6] Korea Software Copyright Committee, Patent Troll and SW patent, Science Information, Korea, 2007.

[7] Y.G. Lee, "Technological convergence and open innovation in the mobile telecommunication industry," Asian Journal of Technology Innovation, vol.16, no.1, pp.45–62, 2008.

[8] C.Y. Wong, "Modeling the dynamics of science and technology diffusion of selected Asian countries using a logistic growth function," Asian Journal of Technology Innovation, vol.17, no.1, pp.75–100, 2009.

[9] Y.J. Lee, "Characteristic features of valuable patents: The difference between private firms and public research institutes in Korea," Asian Journal of Technology Innovation, vol.16, no.1, pp.187–210, 2008.

[10] D. Guellec and B. van Pottelsverghe de la Potterie, "The internationalization of technology analyzed with patent data," Research Policy, vol.30, no.8, pp.1253–1266, 2001.

[11] S. Kutznets, "Innovation activity: Problems of definition and measurement," in The rate and direction of inventive activity, ed. R. Nelson, Princeton University Press, Princeton, NJ, 1962.

[12] M. Hirschey and V.J. Richardson, "Valuation effects of patent quality: A comparison for Japanese and U.S. firms," Pacific-Basin Finance Journal, vol.9, no.1, pp.65–82, 2001.

[13] D.C. Mowery, J.E. Oxley, and B.S. Silverman, "Technological overlap and interfirm cooperation: Implications for the resource-based view of the firm," Research Policy, vol.27, no.5, pp.507–523, 1998.

[14] R.J.W. Tijssen, "Global and domestic utilization of industrial relevant science: Patent citation analysis of science-technology interactions and knowledge flows," Research Policy, vol.30, no.1, pp.35–54, 2001.

[15] B.U. Yoon and Y.T. Park, "A text-mining-based patent network: Analytical tool for high-technology trend," Journal of High Technology Management Research, vol.15, pp.37–50, 2004.

[16] W.H. Lee, "Technology innovation in Korea through patent citation analysis," KIIE, vol.1, pp.1007–1013, 2003.

[17] S.H. Yoo, "A study on estimation of technology life span using analysis of patent citation," Information Management Research, vol.35, pp.93–112, 2006.

[18] Y.H. Tseng, "Text mining techniques for patent analysis," Information Processing & Management, vol.43, pp.1216–1247, 2007.

[19] J.O. Lanjouw and M. Schankerman, "Patent quality and research productivity: Measuring innovation with multiple indicators," The Economic Journal, vol.114, no.495, pp.441–465, 2004.

[20] B.-W. Lin, C.-J. Chen, and H.-L. Wu, "Predicting citations to biotechnology patents based on the information from the patent documents," International Journal of Technology Management, vol.40, no.1, pp.87–100, 2007.

[21] W.O. Christian and C.L. Giles, "Extraction of rules from discrete-time recurrent neural networks," Neural Networks, vol.9, no.1, pp.41–52, 1996.

[22] T.G. Biing and Z.G. George, "An artificial neural network with partitionable outputs," Computers and Electronics in Agriculture, vol.16, no.1, pp.39–46, 1996.

[23] M. Fattori, G. Pedrazzi, and R. Turra, "Text mining applied to patent mapping: A practical business case," World Patent Information, vol.25, pp.335–342, 2003.

[24] C. Clifton and R. Cooley, "TopCat: Data mining for topic identification in a text corpus," Proc. Third European Conference of Principles and Practice of Knowledge Discovery in Databases, Prague, Czech Republic, 1999.

[25] K.-K. Lai and S.-J. Wu, "Using the patent co-citation approach to establish a new patent classification system," Information Processing & Management, vol.41, no.2, pp.313–330, 2005.

[26] O. Coskunoglu, B. Hansotia, and S. Muzaffar, "A new logit model for decision making and its applications," The Journal of the Operations Research Society, vol.36, no.1, pp.35–41, 1985.

[27] R.A. Fisher, "The use of multiple measurements in taxonomic problems," Annals of Eugenics, vol.7, pp.179–188, 1936.

Won-Gyo Jung received the M.S. degree in information management engineering from Korea University in 2009, and the B.S. degree in industrial engineering from Kyunghee University in 2007. He is now a doctoral candidate of information management engineering at Korea University. His research interests are information system, e-business and data mining.

Sang-Sung Park received the M.S. and Ph.D. degrees in industrial engineering from Korea University in 2003 and 2006, respectively. He is now a research Professor of Division of Information Management Engineering at Korea University. He teaches strategic management of electronic business, computer system, and strategic application of Patent information. His research interests are pattern recognition, data mining, patent analysis.

Dong-Sik Jang received the Ph.D. degree in industrial and systems engineering from the Dwight Look College of Engineering, Texas A&M University, in 1988, the M.S. degree in Industrial Engineering from The University of Texas in 1985, and the B.S. degree in Industrial Engineering from Korea University, in 1979. He is now a professor of information management engineering at Korea University in Seoul, Korea.