Towards an Efficient R&D Theme Prediction with Machine Learning

Masashi Shibata[1], Koichi Inoue[2], Yuichi Ohtsuka[2], Kazuhiro Fukuyo[2], Masakazu Takahashi[2]
[1]Fujitsu Kyushu Network Technologies Limited, Fukuoka Japan
[2]Graduate School of Innovation and Technology Management, Yamaguchi University, Japan

Abstract--**This paper proposes an efficient method for R&D theme selection. There are various methods for the R&D theme selection such as patent analysis, paper survey and delphi investigation and so on. Patents and peer reviewed papers are easy to obtain, and are frequently used as materials for the selection method. In addition, making use of the national projects such as the social infrastructures feasibility project is one of the efficient theme selection method. A survey shows that the R&D theme selection is one of the biggest challenge for the private company. Generally, short term R&D theme selection is aiming at implementation within 2 years such as commercialization and sales expansion. Long-term theme is used for the R&D investment budget plan. On the other hand, Medium-term R&D theme is often aimed implementation within 5 years such as an exploratory technology theme. Since it relies on the heuristics knowledge for the theme selection with technology trends, an efficient selection method of the medium-term theme is required. Many technological analyses with the intellectual properties are performed so far. In this paper we propose a method of selecting the R&D theme using a machine learning based on public information.**

I. INTRODUCTION

In the corporate activities, it is very important to understand the future technology trends to perform a variety of management decision making. There are many methods to analyze the trends such as the patent information analysis, the paper survey and so on. Among the analysis methods, the information analysis from the patents is mainly used for the technology trend analysis in Japan. This is because the information source is open to the public, and it is easy to get from the Web site. Furthermore, the patent information has following characteristics; a) Due to the patent submission is originally industrialization basis, easy to understand the technological taste, b) It has rich information in not only the patent body but the classification codes. For these reasons, many patent analyses are made for understanding the technology trends in the business sector.

A typical patent analysis method makes use of the patent database such as the J-PlatPat [1]. From the database. Analysts extract the target patents based on keywords along with the research purpose and make the mapped patents from the results of heuristics analysis. Furthermore, the NTCIR (NII Testbeds and Community for Information access Research) project proposed the method for the automatic classification and patent search with text mining or machine learning [2].

Thus, the patent information is one of the useful data sources in order for forecasting the technology trends. However, the analytical method focusing on the patent document has some challenges. It is difficult to apply common text mining method for the unique wording and the long sentences. Since the chart has an important meaning, only the document information is not enough to extract the content sufficiently. Therefore, the patent analysis is performed manually and its results are heuristic so far.

The Link mining method is one of the techniques to visualize the relationship structure of things [3]. This method is mainly used for the relationship analysis for such networks as web pages, citation, gene network, and so on. Since the conventional method for the analyses depend on the heuristic knowledge, in this paper, we challenge the technological structural analyses and the technological trend forecasting using patent information. Then, we focus on the classification codes in the patents. The technological relations are visualized from the graph made by the classification codes. Then, the features of the graph is calculated by the link mining method. Then, the future technological relations are predicted by the machine learning method using the features.

The rest of the paper is organized as follows: Section 2 discusses the research background and related work; Section 3 briefly summarizes the gathered data on the target patent sector; Section 4 describes the analytical method of the data; Section 5 presents analytical results; and Section 6 gives some concluding remarks and future work.

II. RELATED WORKS

First, we describe the analytical methods on the business sector. Many technological analyses with the patent information were made for the business solution so far [4-9]. They also tried to extract the solution hitting for their problems. A theory of the TRIZ (Teoriya Resheniya Izobretatelskikh Zadatch) aims for the problem solving from the inconsistencies in the patent [10-13].

Kawakami et al, proposes idea generation support system from the TRIZ concept to design system making use of the FUBEN-EKI (FUrther BENEfit of a Kind of Inconvenience). They proposed the contradiction matrix and make use of the matrix for the creative thinking [14].

The KT (Kepner-Tregoe) method analysis is focusing on the problem solution among the business sector. In this method, human pattern of thinking, reaction, and activity are classified from four questions; Situation appraisal, Problem analysis, Decision analysis, and Potential problem analysis. Primary benefits of the KT method is factor analysis of the problem. On the other hand, but, the expertise is required to generate the knowledge [15, 16].

Then, we focus on the patent analysis. One of the techniques of the patent information analysis is the patent mapping. Kiriyama made content analyses with this method [17]. Shide et al, were performed using the patents analyses the change of the positioning for customer of research and development activities of the company [18]. Kimura proposed the technology evaluation method based on the patent analysis for technological strategy planning [19].

As for the analysis of important information derived from the patent analyses, Carpenter analyzed the important cited patents. Muguruma shows the validity of the patent citation analysis to propose the FCA (Forward Citation Applicant) map [20, 21]. Sato et al,, proposed the importance calculation method of the patent document based on the citation information [22].

Ogawa et al, proposed a basic patent extraction based on the citation information [23]. Albert conducted a validation of citation for the important patent among the industry [24].

Instead of the heuristic knowledge extraction, the automatic classification of the patent is proposed. Tanaka proposed a method of extracting the patent characteristics automatically [25]. Yamashita proposed a method of surveillance technology and specific method of patent classification with text mining [26]. Yamamoto et al, proposed a method to enhance the compatibility of the search by applying the information of the related patent documents in search of the academic papers. Yamamoto, proposed a method to find the scientific papers with a variety of the further information [27, 28].

Kleinberg extracted the topic and the description of the relationship with the graph theory [29]. Eto proposed a measure of the co-citation based on the structural units of the paper [30]. Ueda proposed a technical analysis with the active mining method focuses on the cognitive processes of the patent examiner, making use of the patent classifications [31].

Thus, with the application of the information technology, Knowledge extraction from the patent is performed. Furthermore, Okumura et al, performed the recall and accuracy improvement of the patent search with the natural language processing and summarizing automation of patent classification [32]. Okamoto et al, proposed a method to extract the task for the research and development applying the analogy of the short-term memory recall in the human brain. They resembling the citation of the entire patent on a large scale network [33]. Inoue et al, made a corporate network that made the joint application of the patent. From the results of the analysis, they found the network had a scale-free property. Then, they extracted the new patent network with the geographic characteristics [34].

Ogawa et al, extracted the potential research domain with analyzing the article and the patent information. The references in the articles and the patents were made to generate the cluster includes the research in the same domain. The average publication year and the number of the articles in each cluster provides the useful information to comprehend the emerging research fields. Furthermore, the keyword similarity of the scientific article cluster with the patent was calculated to extract the potential and non-commercialized research area [35]

Thus, the patent analysis was carried out in various ways, especially the text mining and the machine learning are used for such as the patent search automation and the improvement of the recall and the precision for the patent search. It is also used the link mining method for the analysis of the citation relationship and the joint application relationship among the patents.

However, the studies for the link mining methods to analyze the classification codes such as the theme code and the F-term are unexplored so far. Thus, in this paper, the technological structure of the patent is visualized by analyzing the classification codes by link mining method. With our proposed method, the technological relation constitute of patent group. Then, the transition of the relations is predicted as a result.

III. EXPERIMENTAL CONFIGURATION

The patent information consists of the metadata, the literal information, and the graphical information. The Japanese patent classification system has the three Meta data such as the IPC (International Patent Classification), the FI (File Index) and the F-term (File Forming Term). Both the FI and the F-term are used to indicate the patent technological field more detail than the IPC. The F-term is further divided into the theme code and the viewpoint.

We focus on the seven technological fields: 1) the Information transfer between computers, 2) the computer and the data communications, 3) the telephonic communication services, 4) the data exchanges in the wide-area networks, 5) the small scale networks, 6) the communication control and 7) the mobile radio communication systems, which constitute the core elements of the wide area data exchange technology. Each of these technological field is given its theme code as follows; 5B084, 5B089, 5K201, 5K030, 5K033, 5K034, and 5K067 respectively. We collect the target information from the patent information database by the theme codes [36, 37].

We focus on the three types of patent data; the published unexamined patent application, the published Japanese translations of PCT international publication for patent applications, and the domestic re-publication of PCT international application. They are defined from the submission types as follows; The Published unexamined patent application is the unexamined patent submitted in Japan. The published Japanese translations of PCT international publication for patent applications is the unexamined patent submitted internationally and originally written in the language other than Japanese. The domestic re-publication of PCT international application is the unexamined patent submitted internationally and written in Japanese. The reason why we focus on the three types of the data is for collecting the data not only the domestic data but the overseas submission data.

We gather the data submitted from the year of 1992 to the year of 2011. This is because the electronic data was formed in the Japanese patent system from the year of 1993. Since the application is opened to the public after 18 months from the submission, the data submitted before the year of 1992 is not digitized yet and the data submitted after the year of 2012 are not published at the time of the gathering data. As a result, 251,463 patent data are extracted from the database, and 971 theme codes are given to the patents in total. This means these patents contain 971 technological fields.

Then, we describe the analytical method. A complete graph is made from each patent. This graph is an undirected graph without weight. This represents technological relations within a patent. Theme codes given to the patent are used as the nodes of the graph. Some theme codes are often assigned to a patent for several times since the theme code is a part of the F-term. Regardless of the assigned number of the theme code, the number of the nodes for each theme code is one in the graph. The concurrent number of times was ranked as once to all combinations of the theme codes given to a patent.

The annual graph is made from piling up the complete graphs for each submission year. In this case, the theme code frequency and the theme code co-occurrence are accumulated.

In this paper, we focus on the prediction of the presence or absence of the link between any two nodes, which called the node pair, and cover neither generating nor deleting the nodes in a graph.

Further, it is assumed that the presence or absence of a link of each node pair is determined independently. In this assumption, the link prediction problem is treated as the supervised machine learning problem [38]. The feature vectors of node pairs in known graphs are employed as the training data. The feature vectors of node pairs in unknown graphs are employed as the test data. The presence or absence of links in node pairs are employed as the class label. Then the presence or absence of links in unknown graphs are predicted by them.

Then, we describe the feature vector of the node pair. The eigenvector centrality, the information centrality, and the constraint are used as the elements of the feature vector of the node. Thus it is a three-dimensional feature vector. Eq.1 and Eq.2. are the centrality formula. The constraint shows the volume of the constrained. The constraint formula is given in Eq.3.

$$C_{ev}(i) = \frac{1}{\lambda}\sum_{j=0}^{n} a_{ij} C_{ev}(j) \quad (1)$$

$$C_{inf}(i) = n \Big/ \sum_{j=0}^{n} 1/I_{ij} \quad (2)$$

$$Constraint(i) = \sum_{j \in V_i \setminus \{i\}} \Big(p_{ij} + \sum_{q \in V_i \setminus \{i,j\}} p_{iq} p_{qj}\Big)^2 \quad (3)$$

Where $C_{ev}(i)$ is node-i's eigenvector centrality, $C_{inf}(i)$ is node-i's information centrality, *Constraint(i)* is node-i's constraint. λ shows the maximum eigenvalues of the adjacency matrix, n is the number of the nodes in a graph, a_{ij} shows the elements of the adjacency matrix, I_{ij} indicates information of the vertices. p_{ij} indicates the weight of the vertex.

Eq.4 represents the tensor product of the two nodes' feature vector. $z^{(i,j)}$ denotes the two nodes' feature vector and its direction is node-i to node-j. *x(i)* and *x(j)* denote the feature vector of node-*i* and node-*j* respectively.

Eq.5 represents the symmetrizing tensor products. $\bar{z}^{(i,j)}$ denotes symmetrized feature vector. $z^{(j,i)}$ denotes the two nodes' feature vector and its direction is node-j to node-i.

$$\begin{aligned} \boldsymbol{z}^{(i,j)} &= \boldsymbol{x}^{(i)} \otimes \boldsymbol{x}^{(j)} \\ &= (x_1^{(i)}x_1^{(j)}, x_1^{(i)}x_2^{(j)}, x_1^{(i)}x_3^{(j)}, \\ &\quad x_2^{(i)}x_1^{(j)}, x_2^{(i)}x_2^{(j)}, x_2^{(i)}x_3^{(j)}, \\ &\quad x_3^{(i)}x_1^{(j)}, x_3^{(i)}x_2^{(j)}, x_3^{(i)}x_3^{(j)}) \end{aligned} \quad (4)$$

$$\bar{z}^{(i,j)} = z^{(i,j)} + z^{(j,i)} \quad (5)$$

The co-occurrence of the two nodes feature vector consists of the four co-occurrence relation indices of the two nodes, such as the Common Neighbor coefficient, the Jaccard coefficient, the Cosine coefficient, and the Adamic/Adar coefficient. These coefficients' formulae are given from Eq.6 to Eq.9, respectively.

$$Common\ Neighbor\ coeff = (|I| \cap |J|) \quad (6)$$

$$Jaccard\ coeff = (|I| \cap |J|)/(|I| \cup |J|) \quad (7)$$

$$Cosine\ coeff = (|I| \cap |J|)/\sqrt{|I||J|} \quad (8)$$

$$Adamic/Adar\ coeff = \sum_{z \in |I| \cap |J|} 1/\log(|Z|) \quad (9)$$

Where $|I| \cap |J|$ is the number of times which both node-i and node-j are appeared, $|I| \cup |J|$ is the number of times which either node-i or node-j is appeared, $|I|$ is node-i's appearance number, $|J|$ is node-j's appearance number. z is nodes which have links between both node-i and node-j, and $|Z|$ is node-z's appearance number.

The node pair's feature vector is made by combining the two nodes feature vector and the co-occurrence of the two nodes feature vector. Therefore, it becomes the 13-dimentional vector. If there is a link between the two nodes of the node pair, the node pair's class label value is set to 1, if not, it is set to -1.

SVM (Support Vector Machine) is employed as a classifier. Table.1. shows the parameters of the SVM. There are four parameters such as kernel, C, gamma, and degree. Where kernel is the kernel function of SVM, which has four functions such as liner, rbf (radial basis function), sigmoid, and polynomial. Where C is the penalty parameter, which has the four values. Where gamma is the kernel coefficient, which has the two values. Where degree is the degree of the polynomial kernel function, has the three values. Therefore, there are the 40 parameter combinations in total. The most suitable parameter combination is chosen by the average accuracy score among five trials.

TABLE 1. SVM PARAMETERS

kernel	C	gamma	degree
linear	1,10,100,1000	-	
rbf	1,10,100,1000	0.001,0.0001	-
sigmoid	1,10,100,1000	0.001,0.0001	-
polynomial	1,10,100,1000	0.001,0.0001	3,10,15

Then, the prediction is performed using the selected parameter combination. The prediction results are evaluated by the following indicators such as accuracy, precision, recall, F-value, and AUC (Area Under the Curve). Accuracy indicates the ratio of the correct prediction. Precision indicates the accuracy of the prediction. Recall shows the coverage of the prediction. F value is the harmonic mean of the accuracy and the recall, and is used as an evaluation index of prediction accuracy. AUC is the area under the curve in the plot of the precision against the recall. It indicates the prediction performance by putting the precision and the recall together. These indicators' value are set between 0 and 1, and these are increased in proportion to the predictor's performance.

Table.2. shows the relation of the prediction result and the fact correspondence. Based on Table.2, accuracy, precision, recall, and F-value are calculated from Eq.10 to Eq.13, respectively.

$$Accuracy = (TP+TN)/(TP+FP+TN+FN) \quad (10)$$

$$Precision = TP/(TP+FP) \quad (11)$$

$$Recall = TP/(TP+FN) \quad (12)$$

$$f1 = 2/(1/precision + 1/recall) \quad (13)$$

TABLE.2. PREDICTION-FACT RELATION

	Prediction(1)	Prediction(-1)
Actuality(1)	TruePositive(TP)	FalseNegative(FN)
Actuality(-1)	FalsePositive(FP)	TrueNegative(TN)

Fig.1. shows the conceptual diagram of the precision and the recall. Rectangular area covers the entire prediction. The blue ellipse area covers predicted links exist, Ellipse of dotted line covers the fact has links exist.

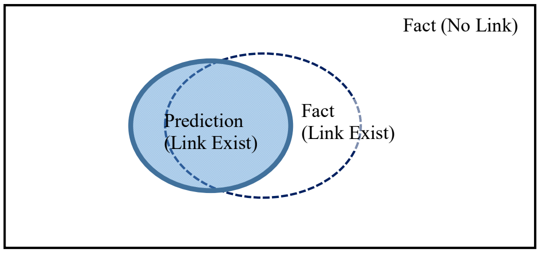


Fig.1 Precision and Recall

We made the combination of the learning data for each year and the class label of the five years later of the learning year and made the data set of each year. The learning period for the prediction is set five years. The learning period is set to five years from the year of 1992 and corresponding class label is set after five years from the learning year. The reason for the setting of the prediction year is as follows; since it takes the eighteen months from the patent application to the open in the Japanese patent system, the prediction term is reasonable to prepare the new business after three-and-a-half years from the open.

Table.3. shows the relation between the year of the learning data, the class label for the learning, the prediction data, and the prediction year. As a result, the prediction term is from 2006 to 2011.

Fig.2 shows the trial flow chart. For the link prediction, the feature vector of the node pair are generated. The feature vector of the node pair has the 13-dimensions, whereas the digit number of the elements is varied. Therefore, the feature vector normalization is required for each element before putting into the classifier.

The standardization procedure is made as follows; a) first standardizing the learning data by removing the mean value and scaling to unit variance on each element, b) standardizing corresponding prediction data using the standardization parameter generated by learning data standardization. c) Both learning and prediction are executed using the standardized feature vectors respectively.

TABLE 3. LEARNING PERIODS AND PREDICTION YEAR

	Learning		Prediction	
	Learning data	Class label	Prediction data	Prediction year
1	1992-1996	1997-2001	2001	2006
2	1993-1997	1998-2002	2002	2007
3	1994-1998	1999-2003	2003	2008
4	1995-1999	2000-2004	2004	2009
5	1996-2000	2001-2005	2005	2010
6	1997-2001	2002-2006	2006	2011

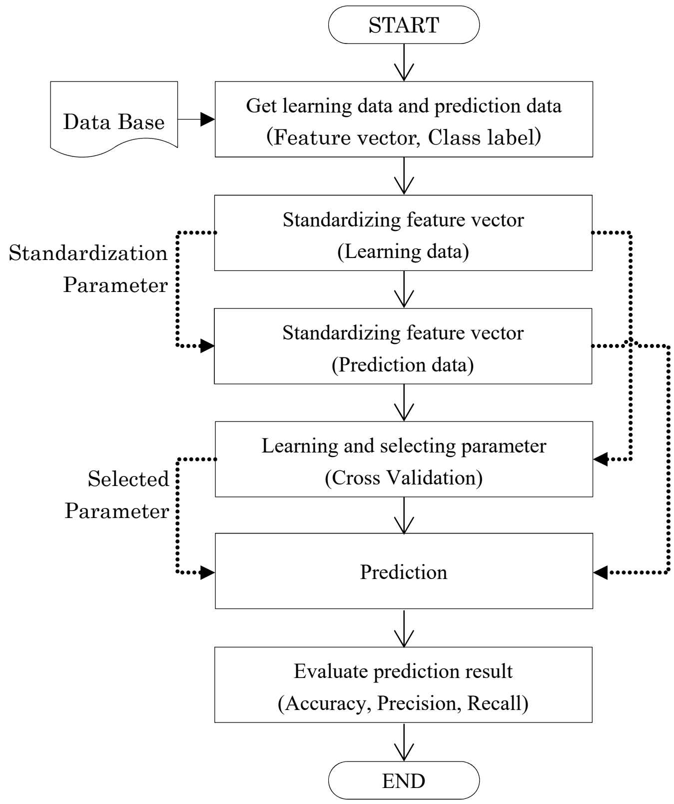


Fig. 2 Trial Flow Chart

IV. EXPERIMENTAL RESULTS

This chapter describes the result of the technological forecasting of the target technological category using the link mining indicators. We focus on 125 out of 971 theme codes, which exists throughout the experiment period. Therefore, the number of node pair is 7,750 because the graphs are undirected.

Table.4. shows the result of the cross validation evaluation in each year. The parameter combination which consists of kernel function = linear, and C = 1, has the highest value of the averaged accuracy over the total period except for 2010. Therefore, prediction is made using these parameters. Due to choosing linear for kernel function, neither gamma nor degree parameter is used.

TABLE.4 SELECTED SVM PARAMETERS

	kernel	C	gamma	degree
2006	linear	1	-	-
2007	linear	1	-	-
2008	linear	1	-	-
2009	linear	1	-	-
2010	linear	100	-	-
2011	linear	1	-	-

TABLE.5 LINK PREDICTION RESULT (COUNT)

	TP	TN	FP	FN	Actuality(1)	Prediction(1)	link ratio
2006	951	5829	493	477	1428	1444	0.184
2007	889	5897	493	471	1360	1382	0.175
2008	770	6117	344	519	1289	1114	0.166
2009	738	6029	624	359	1097	1362	0.142
2010	596	6445	340	369	965	936	0.125
2011	421	6679	238	412	833	659	0.107

TABLE.6 LINK PREDICTION RESULT (INDEX)

	accuracy	precision	recall	f1	AUC
2006	0.875	0.659	0.666	0.662	0.734
2007	0.876	0.643	0.654	0.648	0.725
2008	0.889	0.691	0.597	0.641	0.718
2009	0.873	0.542	0.673	0.600	0.684
2010	0.909	0.637	0.618	0.627	0.691
2011	0.916	0.639	0.505	0.564	0.615
average	0.889	0.635	0.619	0.624	0.695

Table.5. and Table.6. show the result of the prediction. The prediction results are stable throughout the experimental period. The Link ratio of the graphs is from 0.107 in the year of 2011 to 0.184 in the year of 2006. The average link ratio is scored 01.50 in 6 years. This means that the technological area covered by the patents is getting narrowed year by year. The Accuracy values are scored from 0.875 in 2006 to 0.916 in 2011, and the average is scored 0.889 in six years. This means that about the 89% of link conditions are correctly predicted. The Precision values are scored from 0.542 in 2009

to 0.691 in 2008, and the average is scored 0.633 in six years. The Recall values are scored from 0.505 in 2011 to 0.673 in 2009, and the average is scored 0.619 in 6 years. These mean that about the 64% of the predicted links by this predictor are actually generated, however, about the 62% of the actually generated links can be predicted. The F values are scored from 0.564 in 2011 to 0.662 in 2006, and the average is scored 0.624. The AUC value are scored from 0.615 in 2011 to 0.734, and the average is scored 0.695. Fig. 3. shows the precision-recall curve for class label of 2002. Another year of the graphs are also made the same shape.

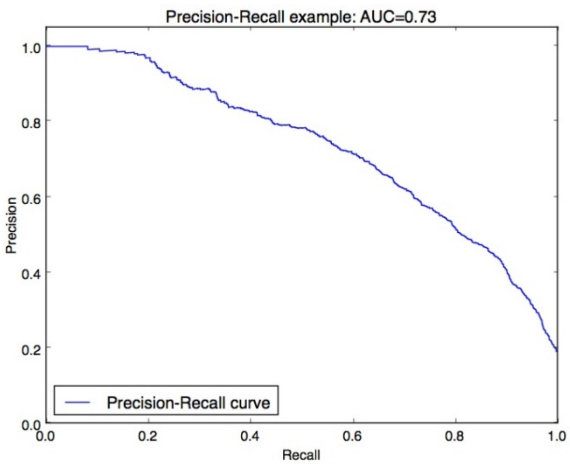
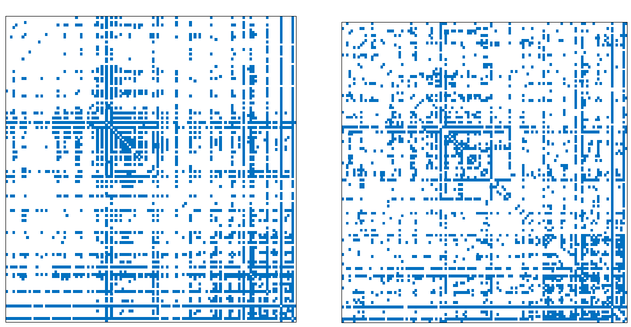


Fig.3. Precision Recall Curve in 2002

The actual technological relations and the predicted technological relations in 2006 are visualized as the adjacent matrixes in Fig.4. (a) and (b) respectively. These sizes are 125x125 and are symmetric due to the undirected graphs. The dots within the graphs represent the relation between 2 technological fields. The predicted graph depicts the density of relations and the technological fields which have many relations with others. We can use this prediction results to forecasting future technological relations.



(a) (b)

Fig.4. Adjacent Matrix in 2002

V. CONCLUDING REMARKS

This paper describes an efficient method for R&D theme selection. We focuses on the Japanese patent data for the future R&D theme prediction with machine learning. We make use of the patent theme codes which constitute the core elements of wide area data exchange technology, indexed 5B084, 5B089, 5K201, 5K030, 5K033, 5K034 or 5K067. This paper, we used SVM as a classifier. As a result, we succeed in predicting the link condition with not only exhaustiveness but stability throughout the period.

Our future work includes improvement both the precision rate and the recall rate with adjusting the parameters and applying to another technology fields for evaluating our proposed method.

REFERENCES

[1] J-PlatPat 2016. [Online]. Available: https://www.j-platpat.inpit.go.jp/web/all/top/BTmTopPage. Accessed: Mar. 25, 2016.

[2] J. I. Lewis and R. H. Wiser, "Fostering a renewable energy technology industry: An international comparison of wind industry policy support mechanisms," *Energy Policy*, vol. 35, no. 3, pp. 1844–1857, Mar. 2007.

[3] Shibata, M., and Takahashi, M., "Toward Technological Trend Forecasting through Geometrical Patent Analyses with Link Mining," Proceedings of the 11th International Conference on Innovation and Management, pp.1070-1076, 17-19 Nov. 2014, Vaasa, Finland.

[4] R. M. Rosenzweig, "The hazards of recombinant DNA," *Trends in Biochemical Sciences*, vol. 2, no. 4, p. N84, Apr. 1977.

[5] H. L. Shuchman, "Engineers who patent: Data from a recent survey of American bench engineers," *World Patent Information*, vol. 5, no. 3, pp. 174–179, Jan. 1983.

[6] F. Narin at el, "Patents as indicators of corporate technological strength," *Research Policy*, vol. 16, no. 2-4, pp. 143–155, Aug. 1987.

[7] V. Oltra and M. Saint Jean, "Variety of technological trajectories in low emission vehicles (LEVs): A patent data analysis," *Journal of Cleaner Production*, vol. 17, no. 2, pp. 201–213, Jan. 2009.

[8] B. Golembiewski, at el, "Identifying trends in battery technologies with regard to electric mobility: Evidence from patenting activities along and across the battery value chain," *Journal of Cleaner Production*, vol. 87, pp. 800–810, Jan. 2015.

[9] M. Karvonen, at el, "Technology competition in the internal combustion engine waste heat recovery: A patent landscape analysis," *Journal of Cleaner Production*, vol. 112, pp. 3735–3743, Jan. 2016.

[10] G. S. Altshuller, *Creativity as an exact science: The theory of the solution of inventive problems*. United Kingdom: Taylor & Francis, 1984.

[11] H. Altov, G. Al'tov, and L. Shulyak], *The art of inventing: How to invent and how to solve technical problems.* Worcester, MA: Technical Innovation Center, 1994.

[12] G. Altshuller, *Innovation algorithm: TRIZ, systematic innovation and technical creativity.* Technical Innovation Ctr, 1999.

[13] G. Altshuller, *40 principles extended edition: Triz keys to technical innovation.* Technical Innovation Ctr, 2005.

[14] H. KAWAKAMI, at el, "Idea Generation Support System for Implementing Benefit of Inconvenience by Employing the Theory of Inventive Problem Solving," *Transactions of the Society of Instrument and Control Engineers,* vol. 49, no. 10, pp. 911–917, 2013.

[15] C. H. Kepner and B. TREGOE, *The rational manager.* Princeton: Princeton Research Press, 1976.

[16] B. B. Tregoe and C. H. Kepner, *The new rational manager: An updated edition for a new world.* United States: Princeton Research Press, 1997.

[17] T. Kiriyama, "IP Information Analysis (<Special feature>Patent Information: Analysis and Effective Utilization)," *The Journal of Information Science and Technology Association,* vol. 60, no. 8, pp.306-312, 2010.

[18] K. SHIDE and M. ANDO, "THE SHIFT OF POSITIONING OF JAPANESE GENERAL CONTRACTORS' R&D ACTIVITIES," *Journal of Architecture and Planning (Transactions of AIJ),* vol. 76, no.

668, pp. 1929–1935, 2011.
[19] H. Kimura, "One Approach of Technology Stocktaking and Evaluation for Corporate Technology Strategies: Emphasizing Future Intentions and Quantification through Patent Analysis," *The Journal of Science Policy and Research Management*, vol. 26, no. 1/2, pp.52-61, 2012.
[20] M. P. Carpenter, F. Narin, and P. Woolf, "Citation rates to technologically important patents," *World Patent Information*, vol. 3, no. 4, pp. 160–163, Oct. 1981.
[21] M. Muguruma, "The Usefulness of Patent Forward Citation Analysis and its Practical Examples," *The Journal of Information Science and Technology Association,* vol. 56, no. 3, pp.114-119, 2006.
[22] Y. SATO and M. IWAYAMA, " A study of patent document score using patent-specific attributes in citation analysis," *Journal of Information Processing and Management,* vol. 51, no. 5, pp. 334–344, 2008.
[23] T. Ogawa, et al, "Finding Basic Patents using Patent Citations," *IPSJ SIG Notes, Information Processing Society of Japan,* vol. 2005, no. 35, pp.41-48, 2005.
[24] M. B. Albert, et al, "Direct validation of citation counts as indicators of industrially important patents," *Research Policy,* vol. 20, no. 3, pp. 251–259, Jun. 1991.
[25] K. Tanaka, "Multi-viewpoint Clustering of Patent Documents," *IPSJ SIG Notes,* vol. 2008, no. 4, pp.9-14, 2008.
[26] Y. YAMASHITA, "Text mining technology for patent analysis and patent search: Patent search and patent analysis service patent integration," *Journal of Information Processing and Management,* vol. 52, no. 10, pp. 581–591, 2010.
[27] M. Yamamoto, at el, "A journal paper filtering using the profile revised by patent document information," IEEJ Transactions on Electronics, Information and Systems, vol. 130, no. 2, pp. 358–366, 2010.
[28] M. Yamamoto and H. Kinukawa, "Journal paper filtering using multiple information," *Electronics and Communications in Japan,* vol. 95, no. 5, pp. 21–32, Apr. 2012.
[29] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM*, vol. 46, no. 5, pp. 604–632, Sep. 1999.
[30] M. Eto, "A New Co-citation Measure Based on Structures of Citing Papers," Database, 49(SIG 7(TOD 37)), Information Processing Society of Japan, pp.1-15, 2008.
[31] I. Ueda, "Active Mining Utilizing the Patent Classification IPC o F1 o F Term that Focuses on the Cognitive Processes of Patent Examiner," *Proceedings of SIG-FAI*
[32] M. Okumura, et al, *Patent information processing: a natural language processing approach,* (in Japanese), pp.36-47, Corona Publishing, 2012
[33] H. Okamoto, Y. Tsuboshita, "Mesoscopic Analysis of Patent Citation Networks : Towards Enhancing Intellectual Property Rights," (in Japanese), *IPSJ Magazine*, vol. 49, no. 3, pp.306-307, 2008.
[34] H. Inoue, W. Souma, and S. Tamada, "Joint Application Networks of Japanese Patents: Analyses of Geographical Features," (in Japanese), ITEC Working Paper Series, issue. 07-04, pp.9-13, 2007.
[35] T. Ogawa, and Y. Kajikawa, "Extraction of Potential Research Area in Polymer Electrolyte Fuel Cell Technology Using Literature Information," *The Journal of Fuel cell Technology,* vol. 13, no. 1, pp.72-75, 2013.
[36] "YUPASS,". [Online]. Available: http://www.yupass.jp. Accessed: Mar. 25, 2016.
[37] "Japan patent office,". [Online]. Available: https://www.jpo.go.jp/cgi/linke.cgi?url=/torikumi_e/searchportal_e/classification.htm. Accessed: Mar. 25, 2016.
[38] H. Kashima, "Survey of Network Structure Prediction Methods," *Journal of the Japanese Society for Artificial Intelligence,* vol. 22, no. 3, pp.344-351, 2007.