Transdisciplinary Engineering: A Paradigm Shift
C. Chen et al. (Eds.)

doi:10.3233/978-1-61499-779-5-993

Using Machine Learning to Forecast Patent Quality – Take "Vehicle Networking" Industry for Example

Chin-Yuan FAN[a,1], Shu-Hao CHANG[a], Hsin-Yuan CHANG[b], Sung-Shun WENG[c] and Shan LO[c]

[a] *Science&Technology Policy and Information Center, National Applied Research Laboratorie, Taiwan*

[b] *Department of Chains and Franchising Management, Takming University of Science and Technology, Taiwan*

[c] *Department of Information and Finance Management, National Taipei University of Technology, Taiwan*

Abstract. Machine learning has become a key development target globally in recent years. An increasing number of algorithms have been applied to solve practical issues. At the present stage, machine learning technologies have progressed from a pure research topic to tools employed for solving practical issues, becoming a key development direction of practical technologies and a prominent emerging discipline. Furthermore, current machine learning technologies have transformed from tools that supplement decision-making to methods that replace manual decision making when generating optimal decisions. This transformation fundamentally changes the tasks that required relatively long workhours in the past. In addition, this may even facilitate distinctive interpretations to effectively aid researchers and operators in addressing problems from a new perspective. Therefore, this study adopted a machine learning technology, namely artificial neural networks (ANNs), to examine relevant topics in patent quality. To verify the effect and identify the characteristics of machine learning in patent quality analysis, this study focused on the fast-changing internet of vehicles (IoV) industry. tailed analyses of key patents were also performed. Finally, a model of high-quality patents in this industry was developed to serve as a reference for other researchers.

Keywords. Machine learning, Patent quality, Vehicle Networking, internet of vehicles (IoV)

Introduction

Machine learning has become a key development target globally in recent years. An increasing number of algorithms have been applied to solve practical issues. At the present stage, machine learning technologies have progressed from a pure research topic to tools employed for solving practical issues, becoming a key development direction of practical technologies and a prominent emerging discipline. Furthermore, current machine learning technologies have transformed from tools that supplement

[1] Corresponding Author, Mail: cyfan@stpi.narl.org.tw

decision-making to methods that replace manual decision making when generating optimal decisions. This transformation fundamentally changes the tasks that required relatively long workhours in the past. In addition, this may even facilitate distinctive interpretations to effectively aid researchers and operators in addressing problems from a new perspective.

Therefore, this study adopted a machine learning technology, namely artificial neural networks (ANNs), to examine relevant topics in patent quality. In patent quality determination, previous practice has mostly depended on experts giving appropriate interpretations on the basis of patent quality indices. However, obtaining expert perspectives and judgements is extremely time-consuming and entails higher personnel costs. Machine learning can effectively shorten the time for obtaining interpretations and thereby reduce the cost of making judgements. The tool is highly beneficial for the fast-paced emerging industry of high-tech applications. Fast determination of patent quality can provide research and development and related patent personnel with a rapid grasp of key conditions in the industry, further facilitating research and development personnel to produce specific strategies in response to their competitors' technology planning.

To verify the effect and identify the characteristics of machine learning in patent quality analysis, this study focused on the fast-changing internet of vehicles (IoV) industry. Analyses of the key influencing indices of patent quality in this industry were conducted using machine learning technologies. Detailed analyses of key patents were also performed. Finally, a model of high-quality patents in this industry was developed to serve as a reference for other researchers.

1. Literature Review

Patent quality has been a crucial research topic because improvement of patent quality is essential to industrial and research development. Current evaluations of patent quality are mostly based on patent-related indices or related data on past litigation cases, the latter of which is mainly analyzed through conventional methods. Numerous scholars have analyzed conventional patent-related indices. For example, [1] used social networks to compiled data regarding co-writing works**.** [2**]** adopted numerous factors such as patent family size, forward and backward citations, patent scope, claims, and patent inventors to develop a composite index according to the stability of individual factors for patent quality analysis. [3] performed an in-depth evaluation on the writing quality of patents from the perspective of patent inventors' thought processes. [4] performed an integrated analysis on patent indicators and product life cycles. [5] further researched patent-related indices such as citations, patent family size, patent inventors, and patent age to identify directions for strategic planning. In summary, conventional patent quality analysis mainly involves integrating key indices such as patent citations (forward and backward citations included), patent family size, number of patent claims, and patent inventors. These indicators typically indicate the explicit messages of patents. Employing such messages facilitates the rapid summarizing of patent conditions to produce an overview of patent quality, through which related problems can be identified and addressed. Indices can be regarded as the most prevalent and accessible criteria for patent quality evaluation.

In addition to indices-based analysis, another approach evaluates patent quality by evaluating patent legal status data, which has become a prominent branch of patent

quality evaluation; [6][7][8][9] have evaluated patent quality by using this approach. However, the exact methods used these scholars differed.[6] regarded litigation status as another variable that influences patent quality. By observing patent litigation, [6] assigned weight to this factor in a patent quality model. [7] studied German patent litigation cases from 1993 to 1995 to analyze the feasibility of high-quality patents. [8][9] investigated the effect of multiple variables on patent quality throughout the patent litigation process. The aforementioned five studies have revealed that patent legal statuses (i.e., whether a patent is involved in a lawsuit, patent payment status, and the overall status of patent maintenance) constitute another set of key criteria for patent quality evaluation.

The aforementioned two sets of criteria, namely patent indices and legal status data, remain the core criteria applied in current practices of patent quality judgement. In addition, such judgements are typically supported by conventional economic and quantity models. However, compared with machine learning technologies, these models require more time when making a judgement and involve more complicated conditions. One of the most representative machine learning technologies is ANNs, which were first developed by **[10]** Early ANNs featured a primitive structure and were modeled on the conditions of human neurons. A notable breakthrough in this technology was achieved by Rochester,[11] who created a perceptron network, a model identification-based algorithm that enables 2-layer computer learning using basic addition and subtraction operations. However, the maturation of this system should be attributed to[12] who developed the backpropogation algorithm for ANNs.

Rapid progress in networking-capacity enhancement and computing system expansion are occurring in ANN development. Such developments aim at facilitating more complicated and highly intensive judgements. Therefore, this study adopted a hybrid ANN to evaluate key patents in the IoV industry and attempted to effectively improve the capacity of ANNs for patent quality evaluation.

2. Research Methods

The To save considerable time in data interpretation, this study employed a hybrid ANN to interpret IoV-related patent data and identify corresponding high-quality patents.

This study comprised the following steps:

2.1. Searching for IoV-related patents:

The IoV has developed into a prominent industry in recent years. At present, the main development direction in this field is to realize automated driving. Therefore, smart vehicles connected to the IoV would not only be equipped with a driving system of their own, but also have the capacity to manage numerous connections, including vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), vehicle-to-pedestrian (V2P), vehicle-to-cloud (V2C), vehicle-to-home (V2H), and vehicle-to-handheld device (V2HD) connections and data exchanges. These factors are collectively referred to as vehicle-to-everything (V2X) interactions. In light of this concept, smart vehicles are expected to become the medium of a large system that connects the existing possibilities of convenience in transportation and daily-life. Hence, the development of this technology may contribute to upgrading the vehicle, household appliance, internet,

semiconductor, and traffic and transportation industries. This study investigated key patents related to the aforementioned trends of development in key databases worldwide, specifically using the patent data search tool developed by Thomson Innovation to search for patents registered in the United States Patent and Trademark Office. The main keywords used in the inquiries consisted of common terms in IoV research and relevant terms for smart vehicles and existing driver assistance systems (Table 1).

Table 1. Keywords used in patent search.

Key words
V2X (Vehicle To Everything)
V2V(Vehicle to Vehicle),
VANET(Vehicular ad hoc network)
V2P (Vehicle to Pedestrian),
V2I (Vehicle to Infrastructure),
V2C (Vehicle to Cloud)
V2H (Vehicle to Home)
M2M(Machine to Machine)
Automotive navigation system
WAVE, Wireless access for vehicular environment
Fleet telematics system
Intelligent vehicle technologies
DSRC, Dedicated short-range communications
IVI , In-vehicle information system
ADAS (Advanced Driver Assistance System)
Parking Assist(Aid) System
Backup parking aid system
Night vision system
LWDS, Lane departure warning system
Blind spot detection system
AFS, Adaptive front-lighting system)
ACC, Adaptive cruise control
CMS, Collision mitigation system
TPMS, Tire-pressure monitoring system

After related patent keywords were determined, the search scope and time were configured. Patents granted by the United States Patent and Trademark Office during 2005–2016 were searched to identify those containing the aforementioned keywords in their title, summary, or right claims. A total of 11,335 patents were filtered, from which patents related according to the INPADOC patent families were excluded (for patents with overlapping technologies, the INPADOC database groups them in families and

present only the most representative entry as a reference). Finally, 4,683 patents were retained as the analysis subject in this study.

Key patent indices that were employed to examine the aforementioned patent data in all sections were then integrated and are presented in Table 2.

Table 2. Key patent indices.

Index	Definition
Claim count	The claims describe the technologies related to a patent. Higher counts generally indicate greater level of innovation.
Assignee count	The assignee count calculates the number of patent owners. Higher assignee counts generally indicate that the patents are more likely to be the products of collaboration.
Inventor count	Inventor count calculates the number of inventors who coinvented a patent. Higher inventor counts indicate that more inventors were involved in the invention.
Application to grant day count	The index calculates the number of days between filing a patent application and receiving the patent grant. Fewer days means that the patents were granted at faster speeds.
Priority right to grant day count	The index calculates the number of days between gaining priority rights of a patent and receiving the patent grant. Priority right date specifies the date the first application of a patent is filed. Longer durations mean more comprehensive protection for the patents.
Forward citations	The number of citations the patent has received.
Backward citations	The number of other patents that the patent cited.
Citations to nonpatent literature	Citations to earlier patents and to nonpatent literature
Patent family count	This index analyzes the state of the patent family.
Country count where a specific patent family is granted	This index analyzes the countries where a patent family is granted.
Legal event count	This index analyzes the frequencies of legal events related to a patent, including maintenance fee payment, lawsuits involved, and patent ownership transference.

The key influencing parameters of patent quality were extracted through the aforementioned process of patent indices integration. The most influential parameter, namely legal event count, was then identified and adopted as the principal judging parameter in subsequent analysis.

2.2. Filtering key parameters

A stepwise regression model was adopted to analyze the key parameters identified in the previous step. In this model, the t value and its significance level α were used as referencing indices to determine whether a specific independent variable would be selected. If the $|t|$ test value of an independent regression coefficient was greater than the theoretical t value obtained from t-tables (or if α achieved significance), the

computer system automatically included this independent variable in regression equations; however, if the value did not achieve significance, the computer system automatically excluded this independent variable from regression equations. Analyses were performed on 11 indices, and those retained after the analyses were employed as conditions for developing the judgment equations in the next research step.

2.3. Analysis models

After model computation was completed, one machine learning models, namely self-organizing maps (SOMs; also known as a Kohonen map) were employed to group related data sets. SOMs is one kinds of grouping algorithm (unsupervised algorithm) based on an ANN. In contrast to other grouping algorithms, SOMs feature a topological map in which the distributions of all outputs (clusters) can be presented. Hence, SOMs present the original high-dimensional data visually in a low-dimensional space, effectively displaying the grouping results.

After the completion of data collection, SOM was used to examine the classification and grouping results of IoV data. Professional research personnel were then invited to analyze and summarize the patent classification results. Finally, an appropriate conclusion was reached.

2.4. Building quality evaluation models

After the aforementioned models were built and achieved stability, each patent was categorized in its corresponding quality class. A regression equation was developed for patents in each group to facilitate evaluating the achievements of all patents in the specific patent groups. Subsequently, the indices and their weights that were employed in individual group evaluations were integrated to assist experts with further interpretation.

2.5. Analysis results of IoV-related patent quality

After the patent quality-evaluating model for IoV-related patents was established, the model was analyzed. On the basis of analysis and research results, related possibilities were investigated and conclusions were drawn.

3. Experimential result

In this paper, we collected patent data (4683 items) on the Vehicle Networking industry from the patent database of Thomson Innovation (https://www.thomsoninnovation.com/login); we split these patent data into five parts. Four parts are the training data set and the other is the testing data set. In addition, each patent data includes 71 patent indicators, and we selected 11 patent quality indicators in our research, as shown in Table 2.

We used “Legal event count” to define the quality of the patent. The model considers the remaining indicators to be features of the patent. As mentioned in Section 2, patent litigation can be used to measure patent quality. Therefore, “Legal event count” is used to define the quality of patents in this research. We define a patent with the value of 'INPADOC Legal Status' from 0 to 1 as Low-Quality; from 2 to 4 as Medium-Quality; and larger than 5 as High-Quality.

Following SOM step (please see table 3), this research shows cluster 4 is the best result in experimential, and through this step , we can define that all this patent from 4groups, low quality patent , medium-low quality patent , medium quality patent , and high quality patent. All this item shown in Figure 1.

Table 3. SOM Cluster Result.

SOM Cluster Process
SOM mode : online
SOM type : numeric
Affectation type : standard
Grid : Self-Organizing Map structure
Features : topology : square x dimension : 10 y dimension : 10 distance type: euclidean Number of iterations : 23065 Number of intermediate backups : 5 Initializing prototypes method : random Data pre-processing type : unitvar Neighbourhood type : gaussian

Table 4. SOM Cluster Result.

Cluster 3			
Degrees of freedom : 2			
	F	pvalue	significativity
Inventor.Count	1.234	0.29126651	
App..Pub..Date..By.Day.Normal	2293.731	0.00000000	***
Pub..Earliest.Prior.By.Day.Norma	1563.595	0.00000000	***
Count.of.Citing.Patents	5.872	0.00283769	**
DWPI.Count.of.Family.Countries	4674.875	0.00000000	***
NPADOC.Legal.Status.Count	454.749	0.00000000	***

Cluster 4			
Degrees of freedom : 3			
	F	pvalue	significativity
Inventor.Count	0.960	0.41050056	

App..Pub..Date..By.Day.Normal	1724.729	0.00000000	***
Pub..Earliest.Prior.By.Day.Norma	2037.8125	0.00000000	***
Count.of.Citing.Patents	8.299	0.00001672	***
DWPI.Count.of.Family.Countries	5676.984	0.00000000	***
NPADOC.Legal.Status.Count	589.841	0.00000000	***

Cluster 5			
Degrees of freedom : 4			
	F	pvalue	significativity
Inventor.Count	1.228	0.29661496	
App..Pub..Date..By.Day.Normal	1446.085	0.00000000	***
Pub..Earliest.Prior.By.Day.Norma	1682.286	0.00000000	***
Count.of.Citing.Patents	6.795	0.00001893	***
DWPI.Count.of.Family.Countries	4257.989	0.00000000	***
NPADOC.Legal.Status.Count	447.992	0.00000000	***

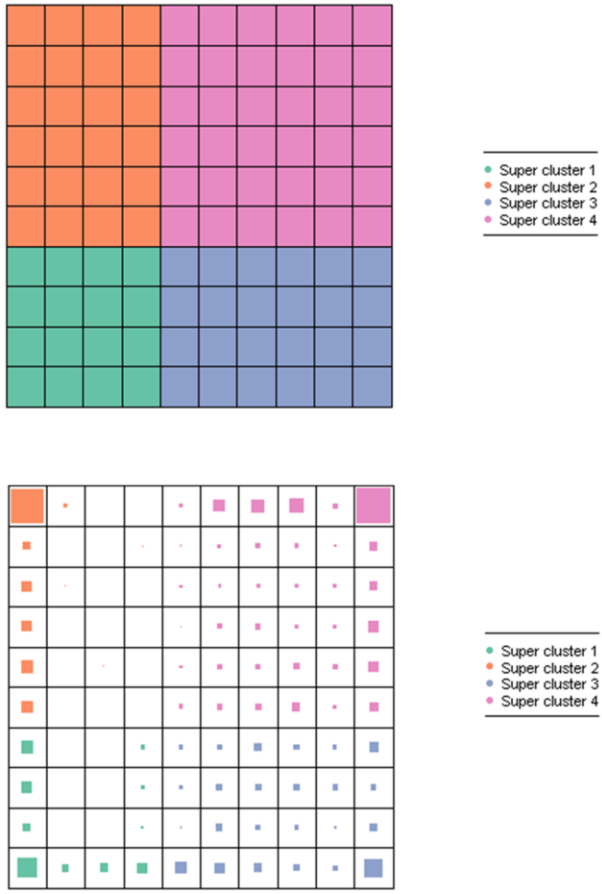

Figure 1. Clustering results.

Following research result, our research combine regression define 4 cluster situation,4683 samples saved at beginning state are used for testing the accuracy of resulting models at this point. The testing outcome are listed below.

1. Prediction function of mid-low value patents:

Y1=-5.003+0.3X1+0.7X2+0.93X3-0.97X4-0.1X5-0.772X6

In mid-low value patents, the significant indicators are the days from application to approval, number of IPC at the moment, the number of inventor, number of non-patent references, number of DWPI families. The rest indicators show no significance.

2. Prediction function of low value patent:

Y2=0.122+0.5X1+0.4X2+0.42X3+0.79X4-0.10X5+0.32X6

In low value patents, just like in mid-low value patents, days from application to approval, number of IPC at the moment, the number of inventor, number of non-patent references, number of DWPI families show significance but the rest are not.

3. Prediction function of mid-high value patents:

Y3=0.64+0.85X1+0.27X2-0.92X3-0.29X4-0.3X5+0.2X6

In mid-high value patents, number of patent inventor is the only significant indicator ,while others are not. Variation of this valuable has an impact on the prediction.

4. Prediction function of high value patent:

Y4=0.47+0.01X1+1.8X2-1.5X3-0.421X4-0. 59X5+0.265X6

In high-value patents, the three significant indicators are number of IPC at the moment, number of patent inventor, and number of CPC while the rest are not significant. That is, changing in these three valuables in the function varies the result most dramatically.

As a conclusion, the only significant valuable is the number of inventor in the high value patent prediction function, both number of non-patent references and number of patent family are important valuables in changing the result of the mid-high value patent prediction function, no significant valuable is find in mid-low patent prediction function, and number of CPC and constant can differ the result of the low value patent prediction function the most.

4. Conclusion

Patents involve complex data which include text, images and lots of metadata; in addition, patents change over time. Therefore, managing organizational knowledge scattered across diverse sources of information is necessary in handling vast data sets. In this research, we collected various patent quality indicators and adapted the SOM to derive these data to develop an automatic patent quality classification model. In this way, when the patent data are updated, the classification model can rapidly re-analyze the patent quality. As we known, patent creation must be publicly discovered in exchange for a time-limited monopoly on its creation. Thus, developing a system that can automatically provide responses when patent data are updated is an important task.

According to the experimental results, the proposed model can obtain the key information from the patent quality indicators. Therefore, the results are very encouraging. It shows that the proposed model can efficiently predict the quality of patents.

Furthermore, further research can examine other data mining techniques for feature selection, such as information gain and principal component analysis to extract the patent quality indicators to enhance the prediction performance.

Acknowledgement

Thanks for R.O.C Ministry of Science and Technology (MOST) support this project, the research project number is MOST 104-2221-E-492-007-MY2.

References

[1] C. Beaudry and A.Schiffauerova, Impacts of collaboration and network indicators on patent quality: The case of Canadian nanotechnology innovation, *European Management Journal,* Vol 29, 2011, Issue 5, pp. 362-376.

[2] M. Squicciarini, H.Dernis and C.Criscuolo, *Measuring Patent Quality: INDICATORS OF TECHNOLOGICAL AND ECONOMIC VALUE*, OECD Science, Technology and Industry Working Papers, 2013.

[3] F. Schettino, A. Sterlacchini and F. Venturinic , Inventive productivity and patent quality: Evidence from Italian inventors, *Journal of Policy Modeling*, Vol 35, Issue 6, November–December 2013, pp. 1043–1056

[4] J.Park and E.Heo, Patent quality determinants based on technology life cycle with special reference to solar-cell technology field, *Maejo International Journal of Science and Technology*; Chiang Mai7.2(May-Aug 2013), pp. 315-328.

[5] M. Grimaldi, L. Cricelli, M. D. Giovanni and F. Rogo, The patent portfolio value analysis: A new framework to leverage patent information for strategic technology planning, *Technological Forecasting and Social Change*, Vol. 94,2015, pp. 286–302.

[6] D. Harhoff, F. M. Schererc and K. Vopeld, Citations, family size, opposition and the value of patent rights, *Research Policy*,Vol. 32, 2003, Issue 8, pp. 1343–1363.

[7] K. Cremers, Determinants of Patent Litigation in Germany, ZEW - Centre for European Economic Research Discussion Paper, 2004, pp. 04-072.

[8] J.R. Allison, M.A. Lemley and J.H.Walker, Extreme Value or Trolls on Top? The Characteristics of the Most Litigated Patents, *University of Pennsylvania Law Review*, Vol. 158, No. 1, December 2009; Stanford Public Law Working Paper No. 1407796. Available at SSRN: https://ssrn.com/abstract=1407796

[9] J.R. Allison, J.H. Walker and M. A.Lemley, Patent Quality and Settlement among Repeat Patent Litigants (September 16, 2010). Stanford Law and Economics Olin Working Paper No. 398. Available at SSRN: https://ssrn.com/abstract=1677785 or http://dx.doi.org/10.2139/ssrn.1677785

[10] W.S. McCulloch and W. Pitts, A Logical Calculus of the Ideas Immanent in Nervous Activity, Bull. Math. Biophysics, 1943, 5, pp. 115–133.

[11] N. Rochester et al. Tests on a cell assembly theory of the action of the brain, using a large digital computer, *IRE Transactions on information Theory*, 1956, 2.3, pp. 80-93.

[12] P.J. Werbos, Experimental implications of the reinterpretation of quantum mechanics, *Il Nuovo Cimento B* (1971-1996), 1975, 29.1, pp. 169-177.