Financial Engineering and Risk Management (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/ferm.2023.060313
ISSN 2523-2576 Vol. 6 Num. 3

Analysis of Key Features of High-Value Patent Based on Lasso-Logit

Ge Zhou*, Yuliang Tong, Haiyou Wang

State Grid Economics and Technology Research Institute Co. Ltd, Beijing, China
**Corresponding author*

Keywords: High-value patent, china patent award, evaluation index, lasso-logit model

Abstract: High-value patent is the key factor for improving corporate core competitiveness, which carries great significance for achieving high quality corporate development strategy. The "14th Five-Year Plan for the National Protection and Application of Intellectual Property Rights" issued by the State Council clearly states the need to shift from focus on quantity to quality development in advancing intellectual property related work, thus enabling high-quality development of intellectual property rights. In order to provide scientific and effective evaluation of high-value patents, this paper centres on the connotation of high-value patents, analyzes 914 samples including control patents and patents that won gold and silver prize in 15th to 23rd China Patent Award. According to the formation stage of high-value patents-- technical invention, application review, authorization protection and market transformation, a high-value patent evaluation index system consisting of 15 related indexes was constructed. Key features of high-value patents were identified and screened based on Lasso-logit method in order to provide index references for the identification and cultivation of high-value patents.

1. Introduction

Innovation is the primary driving force for development, and intellectual property plays an even more prominent role as a strategic resource for national development and a core element of international competitiveness [1]. As the core element of intellectual property, patent provides institutional guarantee to promote and protect knowledge innovation, advance and guarantee industrial development. According to the "World Intellectual Property Indexes 2022", China received 1.59 million patent applications in 2021, more than twice the number received by the United States. The number of patent applications in China presents an explosive growth trend (Figure 1). Figure 2 shows that, the number of patents granted in China maintains a fast growth rate, showing rapid development in 2019-2021. The "14th Five-Year Plan for the National Protection and Application of Intellectual Property Rights" clearly states the need to shift from focus on quantity to quality development in advancing intellectual property related work, thus enabling high-quality creation, efficient use, high-standard protection and high-quality services in intellectual property rights. By leading China's innovation capacity and patent creativity, the intellectual property strategy significantly enhances patent quantity and quality. However, prominent problems still plague China's intellectual property construction, including inadequate development of high-

value patents, loose patent protection, and poor application efficiency. It is necessary to shift from patent quantity rise to quality improvement in our intellectual property construction, which is the inevitable path for China to build intellectual property power.

When China actively formulates the macro-planning for the development of high-value patents, predecessors analyzed the key features of high-value patents based on theoretical and empirical research, accurately evaluated patent value to optimize the high-value patent system. High-value patents refer to patents with high economic value, high market value, high technical value and high legal value [2,3]. The analysis on key features of high-value patents means the theoretical or empirical analysis of important factors in patent value and their correlation with patent value by using patent information from patent database [4]. Some scholars identify the features of high-value patents from the perspective of market value, and based on economic data, use few indexes to evaluate the value of patents with high conversion profits and good market prospects in hot technical fields [5,6]. Although market-based approach sufficiently captures market information, it is often time-consuming and laborious, with strong subjectivity [1]. Secondly, from the perspective of technology, the advanced nature of patented technology is evaluated based on indexes of patent citation, citation status, patent width to highlight patent features of innovation, scientificity and continuity [7]. Finally, from the legal point of view, degree of patent value is measured by indexes indicating legal status of patents, including patent objections, litigation, review, patent length and category, thus reflecting the patent features of risk, stability and validity [8]. In terms of identification way, knowledge such as econometrics, statistics and artificial intelligence is comprehensively used in patent value identification and evaluation model construction, thereby resolving the subjectivity in traditional feature identification of high-value patents [9,10].

Patent value is measured from multiple aspects based on multi-dimensional features. In the case of limited sample information, the preliminary screening of key feature variables plays a prominent role in improving the estimation effect. The Lasso (Least Absolute Shrinkage and Selection Operator) method put forward by Tibshirani combines variable selection and parameter estimation, which demonstrates good variable selection performance in large-scale data variable models and is used in key factor identification. Combined with Logit regression method, it can effectively improve the accuracy in identifying the key features of high-value patents. This study selected 914 patents that won China Patent Award as samples, and examined key features of high-patent value through Lasso-logit method to contribute to the analysis of patent value features.

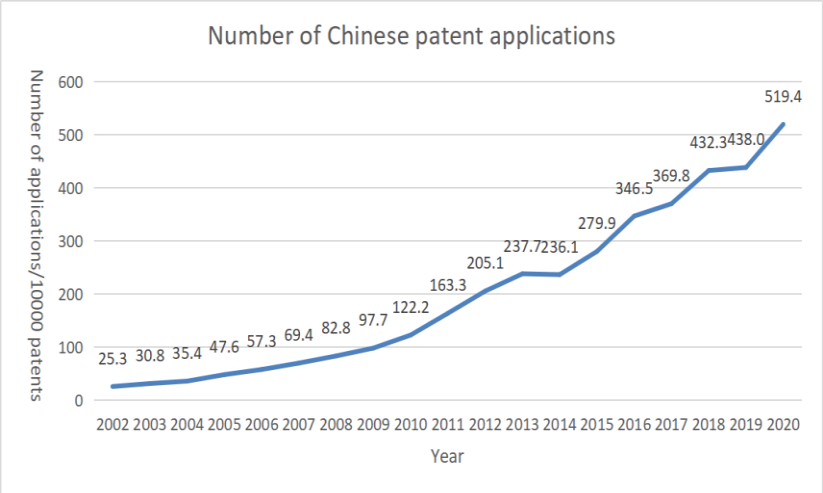

Figure 1: Patent application trend in China

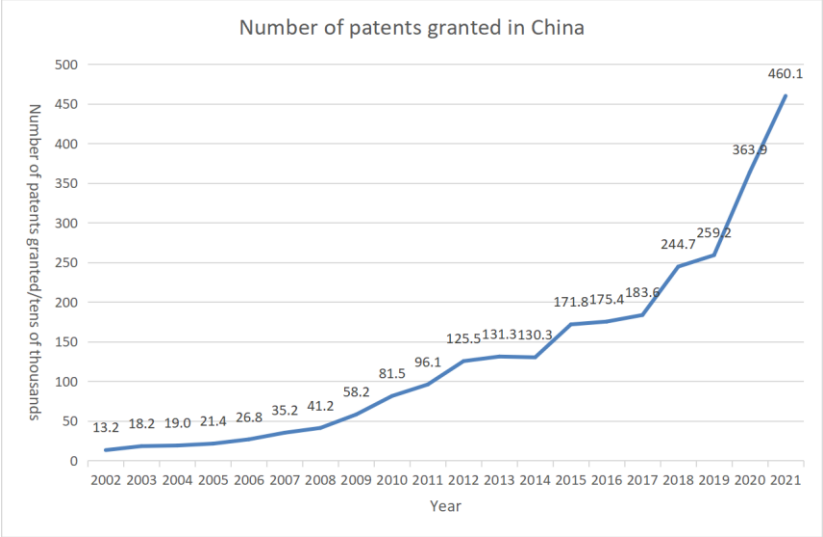

Figure 2: Patent authorization trend in China

2. Research Design

2.1. Definition of Variables and Data Collection

Patent conversion is a complicated systematic process. From idea building, content writing, authorization application, and finally market transformation, it is not only a simple process of knowledge materialization, but also involves enterprises' market income and strategic layout. Hence, the formation process of high-value patents is subdivided into four stages with reference to [2]: technical invention, application review, authorization protection and market transformation. Feature indexes of corresponding stages are selected (as shown in Table 1), and preliminary analysis is made on the feature indexes of each stage.

(1) Technological invention stage: The technological invention stage integrates the formation of creative thinking in the early stage and the breakthrough in key technological points, which also lays the foundation for the overall patent layout and corporate strategic policy of the enterprise in the later stage. Therefore, exploration into the features in technological invention stage helps us understand the key factors in the formation process of high-value patents. As a result, this part incorporates the number of independent claims and the number of inventors into the key feature analysis of high-value patents.

(2) Application review stage: In the application review stage, the patent reviewer conducts preliminary review of the applied patent according to a series of procedures, and then publishes the patent content and the substantial review result. As an indispensible step in the patent formation, it reflects the unique idea of high quality patent, high quality writing, comprehensive patent protection, as well as the strategic layout behind the patentee. Correspondingly, this paper selects the number of subordinate claims, the time limit from application to publicity, the time limit from publicity to authorization, the number of descriptive pages, abstract word count, and whether there is an agency as the key features in this stage.

(3) Authorization protection stage: After the patent authorization, all specific information in the document should be publicized, which involves issues such as citation by patents applied in the later period, definition of the technical field, later maintenance of the patent. Through features such as number of patent citation, legal status, patent duration and IPC category number, it is possible to more clearly understand and discriminate the connotation and value of patents.

(4) Market transformation stage: Mature enterprises need to know how to seize a leading position in the dynamic changing and rapidly developing market. The technical products and method principles in later conversion of patents are the external manifestation of high-value patents. High-value patents can not only help enterprises to seize a leading role among domestic peers, but also consolidate and develop more promising overseas markets. Accordingly, patent life, the number of patent families and patent transfer or not are likely to become key features of high-value patents.

To conclude, this paper selected corresponding factors in the four high value formation stages for analysis. From technology development to the final market application management, the systematic and regular process provides a theoretical basis for the exploration of the key features in the formation of high-value patents.

Table 1: Key features of high-value patents in different stages

patent conversion stage	feature
technological invention stage	Number of independent claims
	Number of inventors
application review stage	Number of subordinate claims
	time limit from application to publicity
	time limit from publicity to authorization
	Number of descriptive pages
	abstract word count
	agency
authorization protection	number of patent citations
	legal status
	patent duration
	IPC category number
Market transformation	expected remaining life
	Patent family
	patent transfer or not

In this paper, a total of 457 invention patents that won gold and silver prize in the 15th to 23rd China Invention Patent Award were collected as research samples. Most invention patents with gold and silver prize demonstrate high quality features of technological novelty, strong originality and great social benefits, which play a leading role in the rating of various indexes. Hence, we selected gold and silver award-winning invention patents as samples of high-value patents.

Meanwhile, uneven data may jeopardise the accuracy of model effect to a certain extent, so stratified sampling method was adopted in this paper. The 457 samples with a number equal to previous Gold and Silver Award-winners in similar technical fields were selected from the 15th to 23rd China Patent Excellence Awards as the control group. The samples were summarized into y=1 (gold and silver award) and y=0 (excellence award) as sample sets for the model construction. y=1 and y=0 had the same proportion to avoid model identification errors from severe proportion imbalance of data.

2.2. Variable Selection Based on LASSO Regression

Lasso regression is to generate an equation of punitive nature through statistical observation of samples. Then, coefficients in the overall model are continuously compressed to avoid over-fitting results, which may incur incorrect conclusions. First proposed by Tibshirani, Lasso regression now

enjoys extensive applications in model prediction [11]. In this paper, the 15 external features of the above analyzed four stages are used as indexes of Lasso model.

Suppose the random sample (x_i,y_i), $i=1,2,...n$; Where, $x_i=(x_{i1}, x_{i2}, ..., x_{im})^T$ is the m-dimensional explanatory variable, and y_i is the explained variable corresponding to the ith observation value. Therefore, when the explained variable y_i is independent and the sample size is known, Lasso estimation herein can be expressed as:

$$(\hat{\alpha},\hat{\beta}) = \arg min\left[\sum_{i=1}^{n}(y_i-\alpha_i-\sum_{j=1}^{m}\beta_j x_{ij})^2\right]$$
$$s.t.\sum_j|\beta_j|\le t \tag{1}$$

Where, t≥0 is a harmonic parameter, the value of t can lower the absolute value of each coefficient in the model. With the continuous decline of t, the regression coefficient gradually approaches 0, or even equals to 0. At this point, the variable with a coefficient of 0 can be deleted.

How to select penalty parameter λ is the key to the model. Commonly used methods include cross validation method, self-help method, etc. In this paper, K-fold cross validation method is selected to determine penalty parameter λ, with specific algorithm shown as follows:

$$C_V(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{\beta}_0^{-k(i)}-x_i^T\hat{\beta}^{-k(i)})^2 \tag{2}$$

Where: $\hat{\beta}_0^{-k(i)}$, $\hat{\beta}^{-k(i)}$ represent the fitting parameter after deleting the kth subset, $k=1,2,...,K$. At this time, the penalty parameter corresponding to the minimum C_V is optimal:

$$\hat{\lambda} = \arg\min_{\lambda} C_V(\lambda) \tag{3}$$

2.3. Establishment of LOGIT Model

Lasso screening makes the key features of each stage more representative. Next, the binary Logit regression model is used for regression analysis of dependent variables. The function P represents the probability for the patent to win gold and silver awards. The above screened features are used as independent variables and denoted as x_1, x_2, … ,x_n to establish the Logit regression model regarding high- value probability of the patent. y is variable 0-1, and a value of 1 indicates that the sample is the gold and silver award-winning patent.

$$P(Y=1) = F(z) = F(x\beta) = \frac{1}{1+e^{(-z)}} = \frac{1}{1+e^{(-tp)}} \tag{4}$$

In the above formula, P (Y=1) represents the probability that 1 is taken as the value of sample t. F(z) is the cumulative probability density function, Y is converted to probability through Logit function, and P is the probability of winning gold and silver awards.

$$\ln(\frac{P}{1-p}) = Z = \beta_0+\beta_i x_i \tag{5}$$

Then, the data of gold and silver award-winning patents and excellence award-winning patents are imported into the model, the model regression results are analysed to obtain the model

coefficients for significance test. Through subsequent analysis, the main features closely correlated with high-value patents are identified.

3. Empirical Results

3.1. LASSO Variable Screening Analysis

Figure 3 shows that, the overall mean square error increases slightly when λ approaches 0.01, but almost no fluctuation occurs in sample error. Hence, this paper believes that this model has a good explanatory power. In Figure 4, when λ is 0.01, the regression coefficient tends to stabilize, and the coefficients of only 9 parameters are not 0, which greatly reduces the model complexity. After comprehensive consideration, λ is optimal when ID is 29, and 9 variables whose coefficients are not zero are screened out, as shown in Table 2.

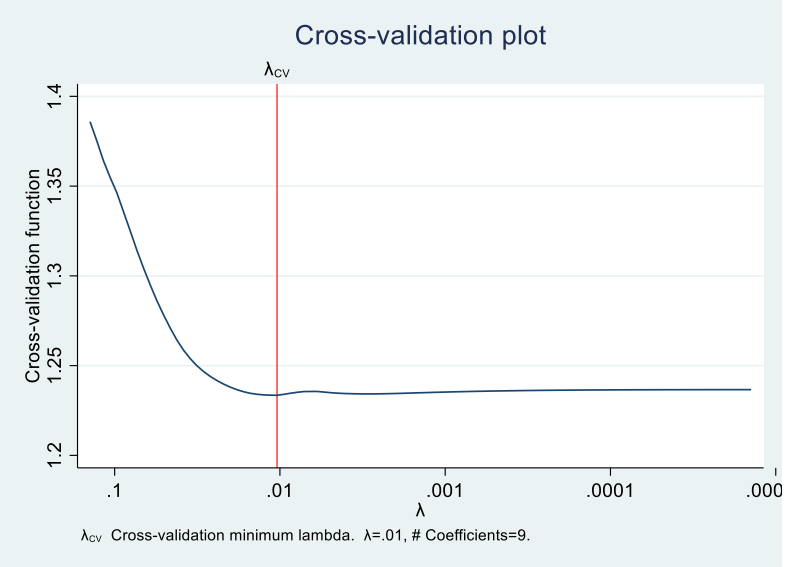

Figure 3: Trajectories of mean square error with λ in cross-validation

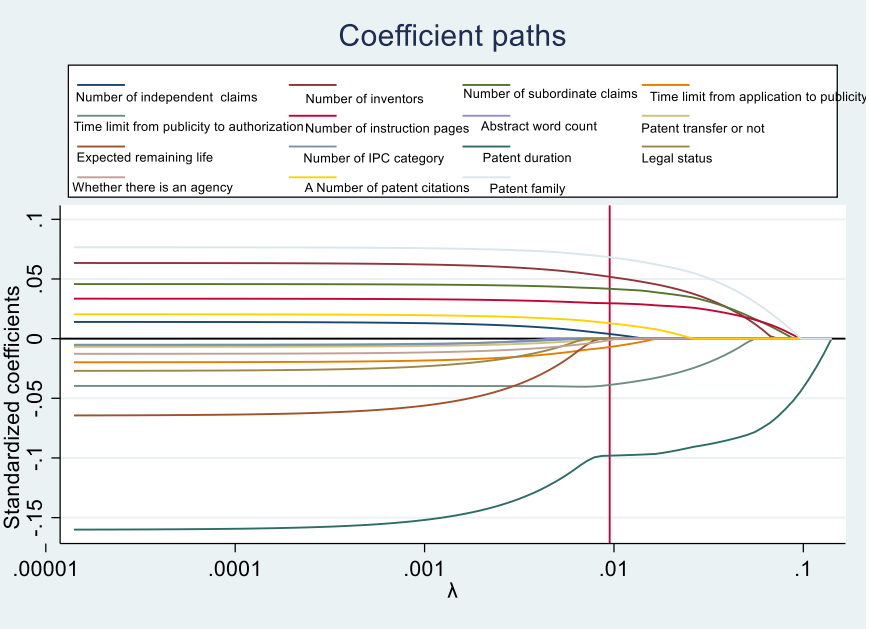

Figure 4: Coefficient path diagram

Table 2: Lasso regression results

ID	lambda	NO. of nonzero coef.	CV mean deviance	
2	0.128	1	1.375	A Patent duration
6	0.088	3	1.336	A Descriptive pages of patent family
7	0.081	4	1.325	A Number of subordinate claims
9	0.067	5	1.304	A Number of inventors
12	0.051	6	1.278	A Time limit from publicity to authorization
19	0.026	7	1.244	A Number of patent citations
24	0.017	8	1.235	A Time limit from application to publicity
29*	0.010	9	1.233	A Whether there is an agency
32	0.008	11	1.235	A Patent transfer or not Expected remaining life
33	0.007	12	1.236	A Number of independent claims
34	0.007	13	1.236	A Legal status
38	0.005	14	1.235	A Abstract word count
40	0.004	15	1.234	A Number of IPC category
100	0.000	15	1.237	U

3.2. LOGIT Regression Analysis

Table 3: Logit regression results

Award-winning patent category	coefficient	Standard error	z value	P value
Number of inventors	0.096	0.023	4.09	0.000
Number of subordinate claims	0.038	0.011	3.30	0.001
Patent duration	-0.131	0.018	-7.06	0.000
Number of patent citations	0.013	0.006	2.19	0.028
Number of patent family	0.123	0.029	4.30	0.000
Constant term	0.168	0.288	0.58	0.560

The probability for the patent to win gold and silver awards is used to judge high-value patents, 9 features screened through Lasso model are taken as independent variables to import the sample group and control group into Stata software. The stepwise regression method is used to build the Logit regression model regarding high-value patents and key features, with the regression results shown in Table 3. According to the regression test results of the model, the confidence level of the variables is significant at the probability of 1%. In this paper, Wald significance test is also implemented for this model, as shown in Table 4. At the significance level of 0.05, the number of inventors, number of subordinated claims, patent duration, number of patent citations and number of patent families all pass the significance test, proving a significant linear relationship between the model indexes and the dependent category variable, which is statistically significant, so parameter analysis and estimation of the model can be further discussed.

(1) The number of inventors, number of claims, number of patent citations and number of patent families have a significant positive relationship with high-value patents. With more inventors, there are more human resources, knowledge resources and technology sources, leading to greater possibility in developing innovative breakthrough technologies. Greater number of claims means greater potential protection scope of the patent, which helps to protect innovative ideas. The number

of patent citations corresponds to the technical correlation between patents, and patents with more citations are usually more important and valuable. Due to the high cost in patent application and subsequent maintenance, enterprises will select patents with higher value for global planning, and the number of patent families reflects the patent value.

Table 4: WALD test

(1)	[Patent award category] Number of inventors =0
(2)	[Patent award category] Number of subordinate claims =0
(3)	[Patent award category] Patent duration =0
(4)	[Patent award category] Number of patent citations =0
(5)	[Patent award category] Number of patent family =0
Chi2 (5) =115.41	
Prob>chi2=0.0000	

(2) Patent duration has a significant negative relationship with high-value patents. Award-winning patents with longer maintenance may lack innovation, so patent reviewer will be more rigorous in judgment of the "three features". Therefore, patents with shorter maintenance are more competitive in the award-winning rating.

4. Conclusion

Based on Lasso-logit model, this paper builds an evaluation index system with the formation stage of high-value patents as the basis to identify and screen the key features of high-value patents. In view of the analysis on the high-value patent connotations and the key features of high-value patents in the formation stage, the corresponding countermeasures are suggested as follows:

(1) Improve the quality of patent application: patent application is a necessary procedure for obtaining patent rights. As the basis of patent quality, patent under application should be novel, creative and practical. Pre-evaluation can be made on patent value and possibility of patent conversion before application to increase the investment and maintenance of potential patents, thus improving the quality of patent application.

(2) Perfect patent review standards: The number of patent applications in China is still in a stage of rapid growth, but simultaneous growth is not achieved in the number and professional literacy of reviewers, resulting in great review workload, insufficient reviewer experience and ability, so some low-quality patents are granted. Relevant units should further perfect the review system, enlarge the patent review team, attach importance to the discrimination and training of high-quality reviewers to constantly improve the patent review quality.

(3) Establish evaluation standard: At present, there is no clear implementation rule for high-value patents in China, and the relevant evaluation standards are not unified, which hinders the cultivation of high-value patents by local governments, enterprises and other innovative subjects. It is necessary to establish a high-value patent evaluation system oriented to government or organization, so that all localities and enterprises can proceed with high-value patent cultivation with reference to the relevant standards.

(4) Optimize patent layout: enhance the frontier and internationality in patent layout, cultivate global consciousness and global vision amid independent innovation, guide more resources into frontier innovation technology, determine the structure distribution and scale of patents in core fields, thus forming a "patent forest".

References

[1] Liu Qin, Yang Rengan, Liu Youhua. High-value patent evaluation methods, existing problems and countermeasures. Science and Technology Management Research, 2022, 42 (04): 147-152.
[2] Chao Rong, Xi Hongzheng. Research on the correlation between the key features of China gold award-winning patents and high-value patents. Information Research, 2019 (10): 8.
[3] Hu Zewen, Zhou Xiji, Ren Ping. Research overview of evaluation and identification of high-value patent based on grounded theory. Information Science, 2022, 40 (02): 183-192. The DOI: 10.13833 / j.i SSN. 1007-7634.2022.02.025.
[4] Zhang Xi, Hu Yuanjia. A Study on the patent value evaluation method without market benchmarking: Theoretical basis, empirical research and future challenges. Soft Science, 2010, 24 (9): 3.
[5] Beaudry C, Schiffauerova A. Impacts of collaboration and network indicators on patent quality: The case of Canadian nanotechnology innovation. European Management Journal, 2011, 29 (5): 362-376.
[6] Caviggioli F, Ughetto E. Buyers in the patent auction market: Opening the black box of patent acquisitions by non-practicing entities. Technological forecasting and social change, 2016 (Mar.): 104.
[7] Ma Yongtao, Zhang Xu, Fu Junying, et al. Review of core patents and their identification methods. Journal of Information, 2014 (5): 38-43.
[8] Hu Caiyan, Wang Xinning. Review of exploration into patent value evaluation methods. China Invention & Patent, 2016 (3): 4.
[9] Bai Limin, Zhu Zhe, Liu Lin, et al. Research on intelligent identification of potential high-value patents in Internet field. China Invention & Patent, 2018, 15 (11): 5.
[10] Li Juan, Li Baoan, Fang Han, et al. Evaluation of invention patent value based on AHP-Entropy weight method: A case study of Toyota open patent. Journal of Information, 2020, 39 (5): 5.
[11] Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B, 1996, 58 (1).