sustainability

MDPI

Article

Prediction of University Patent Transfer Cycle Based on Random Survival Forest

Disha Deng [1,*] and Tao Chen [1,2]

1 School of Management, Wuhan University of Science and Technology, Wuhan 430081, China
2 Industrial Policy and Management Research Center of Hubei, Wuhan 430081, China
* Correspondence: dengdisha@wust.edu.cn

Abstract: Taking the invention patents of the C9 League from 2002 to 2020 as samples, a random survival forest model is established to predict the dynamic time-point of patent transfer cycle. By ranking the variables based on importance, it is found that the countries citing, the non-patent citations and the backward citations have significant impacts on the patent transfer cycle. C-index, Brier score and integrated Brier score are used to measure the discrimination and calibration ability of the four different survival models respectively. It is found that the prediction accuracy of the random survival forest model is higher than that of the Cox proportional risk model, Cox model based on lasso penalty and random forest model. In addition, the survival function and cumulative risk function under the random survival forest are adopted to predict and analyze the individual university patent transfer cycle, which shows that the random survival forest model has good prediction performance and is able to help universities as well as enterprises to identify the patent transfer opportunities effectively, thereby shortening the patent transfer cycle and improving the patent transfer efficiency.

Keywords: random survival forest; patent transfer cycle; cox proportional risk model; Cox model based on lasso penalty; random forest model

check for updates

Citation: Deng, D.; Chen, T. Prediction of University Patent Transfer Cycle Based on Random Survival Forest. *Sustainability* **2023**, *15*, 218. https://doi.org/10.3390/su15010218

Academic Editors: Marco Vivarelli and Antonio Messeni Petruzzelli

Received: 26 October 2022
Revised: 8 December 2022
Accepted: 19 December 2022
Published: 23 December 2022

1. Introduction

Promoting patent industrialization is a key link to facilitate the close integration of science and technology with economy. As a main part of the national innovation system, Chinese universities produce generous patents every year, but the industrialization rate is low and the transfer cycle is long, resulting in a large number of scientific research resources that cannot be fully utilized [1]. Due to the temporality of a patent, if effective technology transfer cannot be carried out within a specific time span, the patent will become invalid, which can cause a huge waste of scientific and technological resources for universities, enterprises and the country [2]. For one thing, as the exclusive right of developed technology, the technology preemption function of patents is essential for the development of enterprises [3]. Clarifying the time from patents' application to authorization can prevent other enterprises from using their patented inventions for commercial purposes, thus reducing competition among enterprises [4]. For another, the cost of maintaining invention patents is relatively expensive for universities, so it is also necessary for universities to sell patents to recover the investment cost of the inventions [5]. Therefore, both universities and enterprises have strong motivations to realize the patent's transformation. Clarifying the fluctuation of patent transfer probability over time can help universities actively seek potential patent grantees in the market at an appropriate time, and then realize patent industrialization. Based on the above reasons, establishing a patent transfer cycle prediction model and analyzing its influencing factors is of great theoretical and practical significance for both universities and enterprises.

The patent transfer cycle, also known as the patent transfer speed and technology time distance, is generally defined as the time difference between the patent application and its transformation. Wang et al. [6] calculated that the technical age of patents ranged from 0 to 20 years from application to licensing, and the average age for Chinese enterprises to choose external technologies was 3.04 years. Lee and Lim [7] concluded that 90% of the patents generated by the Korean public R & D plan could be transferred to the company within 1.61 years based on quantitative analysis. The above researches are not directed towards the patent transfer cycle, but involve this problem in part of the research process, and simply calculate the average time of the patent transfer cycle, granting less consideration to the fluctuation of the patent transfer probability as time progresses. Based on the background above, this paper introduces the random survival forest model into the analysis of an individual university patent transfer cycle, and conducts a dynamic time-point prediction of the patent transfer cycle through modeling, aiming at providing suggestions for both universities and enterprises on identifying the opportunity of patent transformation, thereby shortening the patent transfer cycle and improving the patent transfer efficiency.

2. Background

2.1. Influential Factors of University Patent Transfer Cycle

University patent transfer is a complex process from technology discovery to industrial application. The existing research mainly discusses the influencing factors of patent transfer cycle from the characteristics of patents and the inventor's team.

Firstly, the successful transformation of a university patent depends on the patent's own features. Although universities apply numerous patents, the quality of patents is varied; the patents with high quality are easier to be transferred in the technology trading market [8]. Therefore, patent quality is an important factor affecting the patent transfer cycle. Moreover, due to the territoriality of patent right, patents distributed in different countries can be protected in multiple markets, which can expand their geographical protection scope and reduce legal risk of intellectual property in international business activities for the enterprises [9]; therefore, patents with a larger family size more easily obtain successful transformation and possess a shorter transfer cycle. The successful transfer of university patents also depends on the characteristics of the inventor's team. Previous studies have shown that the participation of university inventors would accelerate the commercialization process of patents [10]. Since patent commercialization activity depends on the social network provided by the inventor, more inventors means more transfer channels [11], which would shorten the patent transfer cycle as well.

To sum up, the research above has generally confirmed the impact of patent features and inventor team characteristics on the university patent transfer cycle, and mainly focused on the factors affecting the university patent transfer cycle and the patent convertibility prediction, granting less consideration to the dynamic effect of patent transfer cycle over time. Nevertheless, when specifically analyzing the issue of university patent transfer, in addition to concern about whether the patent has transferred, the transfer cycle and opportunity of patent transformation also needed to be involved [12]. Thus, this paper aims to incorporate the factors affecting the patent transfer cycle into the survival analysis models based on the existing research, and to predict and evaluate the individual university patent transfer cycle through the model with the best prediction performance.

2.2. Review on the Analysis Method of University Patent Transfer Cycle

Survival analysis was widely adopted by predecessors in the study of patent transfer cycle. For instance, McCarthy and Ruckman (2017) conducted a survival analysis of 54,953 biological patents in the United States from application to licensing using Cox proportional risk model, found that the size of licensors, the number and scope of patent forward citations would shorten the patent licensing cycle, and proposed that further research should be carried out on the patent transfer cycle in subsequent studies [13]. Danish et al. (2020) fitted the technology transfer possibility curve through the Cox proportional risk model, and found that the possibility of patent transfer increased first and then decreased [14]. These survival analysis methods quantitatively analyze the impact of explanatory variables on the patent transfer cycle, but they generally rely on the restrictive assumptions, such as the proportional assumption, and the interaction between covariates are supposed to be taken into account [15]. In addition, the transfer cycle of an individual patent has not been analyzed or predicted in the studies above.

As a derivative method of random forest [16], random survival forest, proposed by Hemant et al. [15], has apparent advantages over other survival analysis methods, especially in high-dimensional data processing. Considering that the university patent samples are large, the characteristic indicators are varied, and the patent transfer cycle does not conform to the normal distribution, which generally contains censored data (whereas the random survival forest model can process the censored data and is superior to the traditional survival methods in determining the nonlinear impact of variables and identifying the interaction of different indicators [15]), random survival forest method is adopted to establish the patent transfer cycle predicting model and obtain the importance ranking of each patent indicator variable. Afterwards, the prediction performances of various survival analysis methods are compared by applying the sample data of university patent transfer.

3. Research Methods

3.1. Survival Analysis

Survival analysis is a statistical analysis method widely applied in varied fields of science, which studies the time until an event of interest occurs. The time span from the starting point to the end point of the event is called survival time.

In survival analysis, survival time is mainly characterized by survival function and risk function. The survival function is defined as:

$$S(t) = P(T > t), 0 < t < \infty \quad (1)$$

It is expressed as the probability that the survival time T exceeds t, where T is a non-negative random variable. The risk function is defined as:

$$\lambda(t) = \lim_{h \to 0^+} P(t \le T < t + h | T \ge t)/h \quad (2)$$

It indicates the instantaneous rate at which the outcome event occurs for the individual which is still alive at time t, that is, the risk of the individual experiencing the outcome event at time t. The risk function is defined as: $\Lambda(t) = \int_0^1 \lambda(s)ds$, the three functions can be converted into each other [17].

In this paper, patent application time is taken as the starting event of the survival analysis and the patent transformation time is taken as the ending event of the survival analysis, with the time interval between the two being the patent transfer cycle. If a patent has not transferred at the end of the observation, it would be regarded as a right censored individual and its transfer cycle would be a right censored data. If the patent has transferred, its transfer cycle would be a complete data.

Based on the existing representative literatures [18–22], it is considered that the Cox proportional risk model [23], Cox proportional risk model with penalty [24], random forest model based on machine learning [25], and random survival forest model [26] are popular in survival analysis research. In addition, since previous studies have shown that the lasso method can obtain better prediction results than the forward and the backward stepwise regression methods [27], the lasso method is adopted to establish a proportional risk model with penalty in this paper. Therefore, Cox proportional risk model, Cox proportional risk based on lasso penalty, random forest model and random survival forest are selected to establish the university patent transfer cycle model, and their predicting accuracy are compared in an effort to select the model with the optimal prediction performance.

(1) Cox proportional risk model

Cox proportional risk model is generally adopted to analyze the influencing factors of individual survival time and to predict the survival or death risk of individuals. Its general form is:

$$h(t|z) = h_0(t)\exp(\beta^T z) \tag{3}$$

wherein, $h_0(t)$ is an unknown benchmark risk function, that is, the risk function when all covariates are taken as 0 or a benchmark value. Equation (3) shows that the benchmark function of death events would be expanded by exp{βx} times under the influence of variable x.

(2) Cox model based on lasso penalty

Lasso method reduces the dimension by punishing the number of regression coefficients to screen out independent variables with much significance, making the model's decision coefficient R^2 larger. The Cox model based on lasso penalty is represented by the minimum of the sum of squares of the residuals plus a penalty function for the regression coefficient, namely:

$$\min_{\beta}\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{P}\beta_j x_{ij}\right)^2, subject\ to \sum_{j=1}^{P}\left|\beta_j\right| \leq \lambda \tag{4}$$

(3) random forest model

The random forest model has high classification accuracy and generalization ability. Random forest algorithm first extracts multiple samples from the original training set samples by means of bootstrap resampling; then, the decision trees are built and combined for each bootstrap sample, and finally the final prediction result is maintained by voting.

(4) random survival forest model

The random survival forest is applicable to the right censored survival data and is the derivative of the random forest algorithm. Its tree building rules are similar to those of the random forest. Firstly, a bootstrap resampling method is adopted to extract multiple training sample sets from the original samples with or without placement, which are recorded as D_i (i = 1, 2, . . . , B). Then, for each sub sample set D_i (i = 1, 2, . . . , B), a model of patent transfer cycle survival tree is established. In the generation process of trees, the criterion of maximizing log rank test statistics is applied to split nodes. The constraint that the number of nodes $d_0 > 0$ is used as the condition for the end of tree growth. To prevent bias, the tree is not pruned after generation. For a survival tree, set $(T_{1,h}, \delta_{1,h})$, . . . $(T_{n(h),h}, \delta_{n(h),h})$ is the information about the patent transfer in the leaf node, where $T_{i,h}$ represents the length of time for the patent transfer, $\delta_{i,h}$ is the dummy variable.

When $\delta_{i,h}$ is taken as 0, $T_{i,h}$ is the right censored data, which indicates that there is no transfer of the patent after $T_{i,h}$. When $\delta_{i,h}$ is taken as 1, it suggests that the transfer occurs at $T_{i,h}$. Calculate the cumulative hazard function of each leaf node in the survival tree as follows:

$$H(t|x_i) = \overset{\Lambda}{H}_h(t) = \sum_{tl,h \leq y}\frac{d_{l,h}}{Y_{l,h}}, x_i \in h \tag{5}$$

where, $d_{l,h}$ represents the number of patents transferred at $t_{l,h}$, $Y_{l,h}$ represents the number of patents not transferred before $t_{l,h}$, $t_{l,h} < t_{2,h} < \ldots < t_{N(h),h}$ represent the discrete time points in leaf node h. Finally, the cumulative hazard function obtained from the random survival forest is:

$$H_e^{*}(t|x_i) = H_h(t) = \frac{1}{B}\sum_{b=1}^{B} H_b^{*}(t|x_i) \tag{6}$$

3.2. Model Evaluation Indicators

Three indicators commonly applied in survival analysis: Brier score, integrated Brier score and consistency index are adopted to evaluate the prediction ability and goodness of fit of the models [28].

(1) Consistency index (C-index)

Harrell's Concordance index (C-index) is applied to measure the global discriminating ability of the model. C-index is independent on selecting a fixed time node to evaluate the model, and particularly takes the situation of individual censored into account. The C-index is defined as follows:

$$CI = \frac{\sum_{i,j\in\Omega} I\left\{\overset{\Lambda}{T_i} < \overset{\Lambda}{T_j}\right\} + 0.5I\left\{\overset{\Lambda}{T_i} = \overset{\Lambda}{T_j}\right\}}{|\Omega|} \tag{7}$$

where I is an indicative function, Ω represents the set of (i, j) (valid pair) legal individuals meeting specific conditions, $\overset{\Lambda}{T_i}, \overset{\Lambda}{T_j}$ and T_i, T_j represent the predicted and actual survival times of individuals i and j, respectively. The prediction error rate of a C-index is generally between 0.5 and 1. A C-index from 0.5 to 0.7 is classified as having low accuracy, 0.7 to 0.9 as having moderate accuracy, and 0.9 to 1.0 as having high accuracy [29].

(2) Brier score and integrated Brier score

In the survival analysis, Brier score (BS) is defined as the mean square of the difference between the observed survival condition and the predicted survival condition. It is an indicator which represents the accuracy of model prediction. Brier score is able to evaluate the model error across multiple time points, and it can be calculated by the individual survival time t, truncated variable δ and sample size N:

$$BS(t) = \frac{1}{N}{}_{i=1}^{N}\{[0 - S(t|x)]^2\frac{I(t_i \le t, \delta_i = 1)}{G(t_i|x)} + [1 - S(t|x)]^2\frac{I(t_i > t)}{G(t|x)} \tag{8}$$

Integrated Brier score (IBS) is the overall measurement of Brier score, which is obtained from the time integration of the Brier score:

$$IBS = {}_{0}^{\max(t)} BS(t)dt \tag{9}$$

Brier score ranges from 0 to 1, where 0 is the best possible value of an applicable model and 0.25 is the highest possible value of an informative model [30]. The smaller the Brier score and integrated Brier score is, the higher the prediction accuracy of the model would be.

4. Data Preprocessing

4.1. Data Source and Indicator System Construction

This paper selects 79,393 invention patents applied for by the C9 League from 2002 to 2020 as research samples. The research observation period is from the patent application time to 1 August 2022. The patent data were obtained from the INCOPAT scientific and technological innovation information platform. The C9 League refers to the nine Chinese first class universities that have signed the Cooperation and Exchange Agreement on Talent Cultivation, including Tsinghua University, Peking University, Zhejiang University, Shanghai Jiao Tong University, Fudan University, Nanjing University, University of Science and Technology of China, Harbin Institute of Technology, and Xi'an Jiaotong University. As a representative of Chinese top universities and an important force of national scientific research output, the scientific research and patent output of the C9 League are of great significance for the technological strength and innovation development of the regions and the country [31]. Consequently, the invention patent application data of the C9 League is taken as the sample to build the university patent transfer cycle model in this paper.

In terms of the indicators' selection of patent transfer cycle, based on the previous research [32,33], 12 specific indicators are selected from three aspects of the patent, which include technology dimension, legal dimension and market dimension, as shown in Table 1.

Table 1. Variables in predictive model.

Dimension	Indicator	Symbol	Explanation	Type
Technical dimension	Technical width	nIPC	Classification numbers	Numerical
	Backward Citation	nBWD_citing	Number of patents citing	Numerical
	Forward Citation	nFWD_citing	Number of patents cited	Numerical
	Countries citing	nBWD_country	Number of countries citing	Numerical
	Countries cited	nFWD_country	Number of countries cited	Numerical
	Non-patent citations	nNPL	Number of non-patent citations	Numerical
	Inventors	nInventor	Number of inventors	Numerical
Legal dimension	Claims	nClaim	Number of claims	Numerical
	Litigation	Litigation	Whether the patent has been sued	Nominal
Market dimension	Family size	Family_size	Number of patents in the same family	Numerical
	Family country	nFamily_country	Number of countries in the same family	Numerical
	PCT application	PCT	Whether the patent is submitted through Patent Cooperation Treaty	Nominal

4.2. Software Realization

During the 20-year observation period, few university patents have been transferred and many more of them have not been transferred. Due to the imbalance of the binary classification outcome variables in the training set, that is, whether the patent has been transferred, it will be difficult for the model to grasp the characteristics of sample indicators, thus reducing the efficiency and prediction accuracy of the model [34]. In order to meet the settings and improve the calculation efficiency of the model, this study applies the SMOTE algorithm proposed by Chawla [35] to solve the problem of unbalanced outcome variables, which is implemented using the smotefamily package of R. Cross validation is implemented using the CoxBoost package of R.

The establishment and evaluation of survival models are realized by R 4.0.1, among which the survival package is applied to establish Cox proportional risk model, glmnet package is applied to establish Cox model based on lasso penalty, ranger package is applied to establish random forest model, randomForestSRC package is applied to establish random forest model, and pec package is applied to compare the models above.

5. Results and Analysis

5.1. Importance of Variables

Since cross validation can be adopted to reduce the deviation and variability of estimation performance, which is caused by the single test set and training set splitting, and ensure that the results obtained do not depend on the random splitting of the selected data [36], this paper compares the prediction error under each parameter selection through 5-fold cross validation. According to the calculation, the number of variables selected at each decision tree node of the random survival forest is 4, the number of variables selected when the node split is 10, there are 200 trees in the forest, and each survival tree has 15 terminal nodes on average. Furthermore, the importance of variables (VIMP), which is defined as the difference between the prediction errors with and without noise, is sorted by calculating the OOB (out-of-bag) error rate. Lager VIMP values indicate variables with high predictive ability, whereas zero or negative values identify non-predictive variables [37]. Calculate the 12 predictive variables under the importance measurement, which is shown in Figure 1.

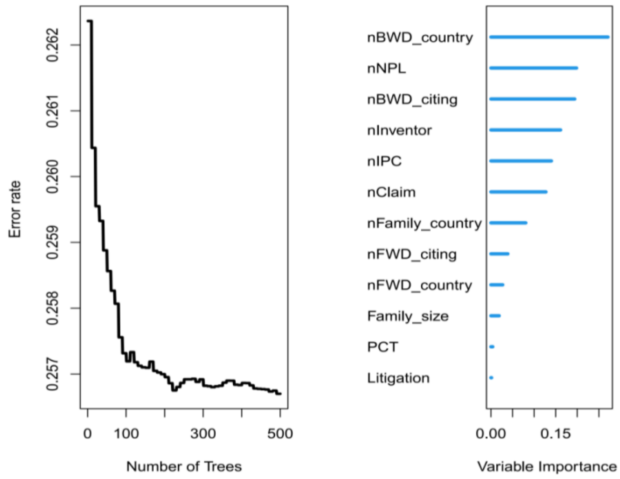


Figure 1. Result of random survival forests.

It can be observed from Figure 1 that the number of countries citing, non-patent citations and backward citations are the three major predictive factors of the patent transfer cycle, which have significant impacts on the prediction results. The number of inventors, technical width, number of claims, number of countries cited, forward citations, and family size are the moderate predictors of patent transfer cycle. Whether the patent is submitted through Patent Cooperation Treaty or whether it has experienced litigation are unimportant predictors, which have little impact on the patent transfer cycle. In order to make full use of the effective information, all variables are taken into the random survival model to predict the results.

Figure 2 shows the overall survival function of the OOB (out-of-bag) patent data, which are generated in the bootstrap process after data preprocessing and test-train data set splitting. The red curve in Figure 2 represents the overall survival rates of the patents, and the green curve represents the Nelson–Aalen estimator, that is, the cumulative incidence of patent transformation calculated by the cumulative risk function, which merely includes the data of patents transferred in the observation period. All the individual university patents in the sample can be predicted through random survival forest, the survival function, cumulative risk function and risk function.

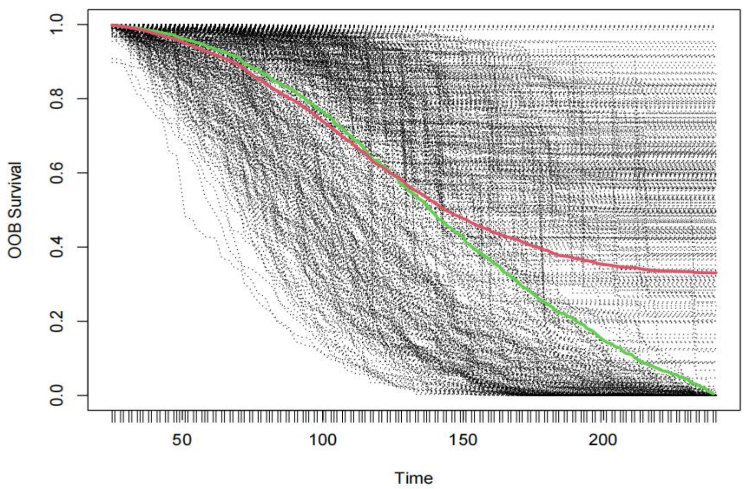


Figure 2. Survival function.

5.2. Model Prediction Comparison

C-index, Brier score and integrated Brier score are applied to evaluate the prediction performance of random survival forest model, Cox proportional risk model, Cox model based on lasso penalty and random forest model. The prediction performance of four models was averaged in the 5-fold cross validation and repeated 100 times. The original data set was randomly divided into training set and test set according to the ratio of 7:3, and C-index is calculated on the train set and test set, respectively. The results are obtained by repeating 100 tests, as shown in Figure 3.

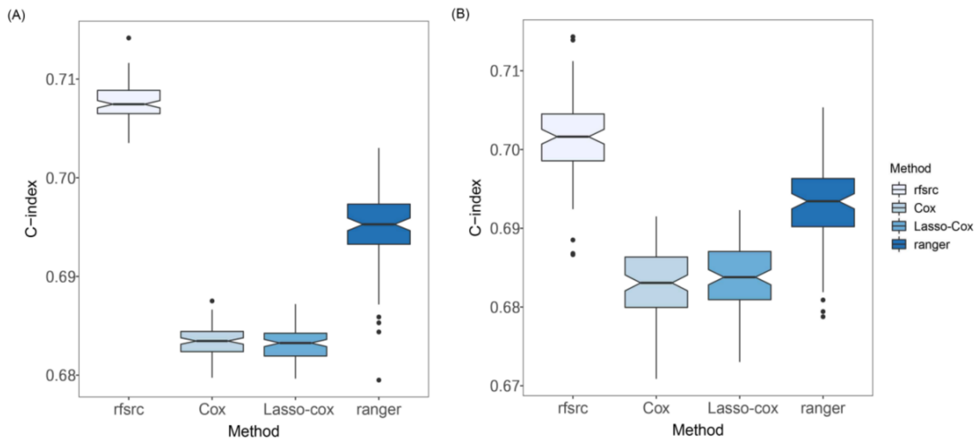


Figure 3. C-index of the four models on the train set (**A**) and test set (**B**).

It can be observed from Figure 3 that the C-index of the random survival forest model on the train set has reached medium accuracy, and the median (0.7075) is larger than that of the random forest model (0.6953), Cox model based on lasso penalty (0.6833) and Cox proportional risk model (0.6835). On the test set, the prediction performance of the random survival forest model is also superior, the median of C-index is 0.7016, higher than that of the random forest model (0.6934), the Cox model based on lasso penalty (0.6838), and the Cox proportional risk model (0.6831).

In order to calculate the Brier scores of the four models, 100 times bootstrap resamples were performed on the original samples. Each bootstrap sample was trained on the data in the bag, the data outside the bag was applied to calculate Brier scores, and finally, obtained the average value of 100 times test. Since the expiry date for an invention patent right is 20 years in China, which means that the maximum time for the patent to transfer is 20 years, that is, 240 months, the time for Brier score calculation is set as 0 to 240. As is shown in Figure 4, the Brier scores of random survival forest model are smaller at most time points than those of the random forest model, the Cox model based on lasso penalty, and the Cox proportional risk model; therefore, the prediction performance of the random survival forest is optimal among the four models.

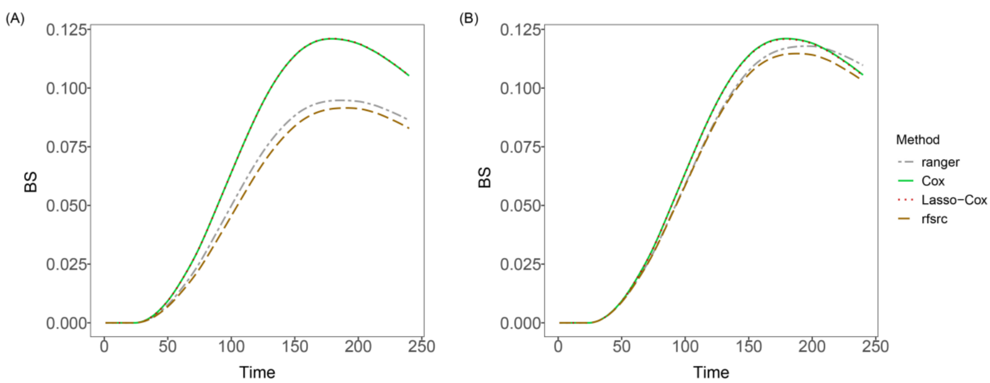


Figure 4. Brier scores of four models on the train set (**A**) and test set (**B**).

In addition, the integrated Brier score is used to summarize the prediction error in the test set, which is shown in Table 2. The integrated Brier score of random survival forest is the lowest among the four models, which indicates the best prediction performance.

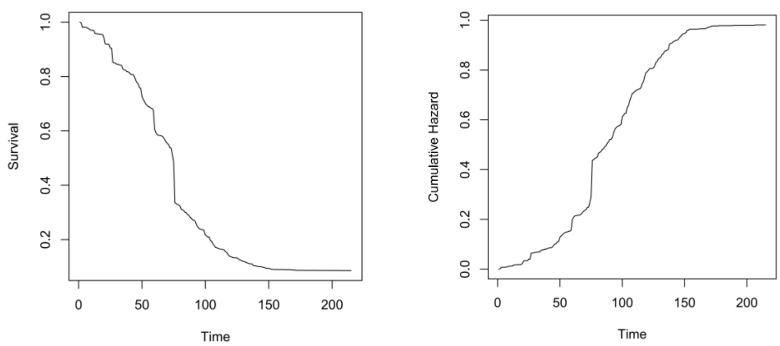
Table 2. Integrated Brier score of 4 models.

	rfsrc	Ranger	Lasso-Cox	Cox
IBS	0.0828	0.0863	0.1053	0.1052

5.3. Case Analysis

To illustrate the predictability of random survival forests in the issue of patent transfer cycle, two representative patent samples applied by Tsinghua University and Zhejiang University are selected for prediction and analysis, which have a large number of patent applications among the C9 League, with the patent applications both exceeding 20% of the research sample.

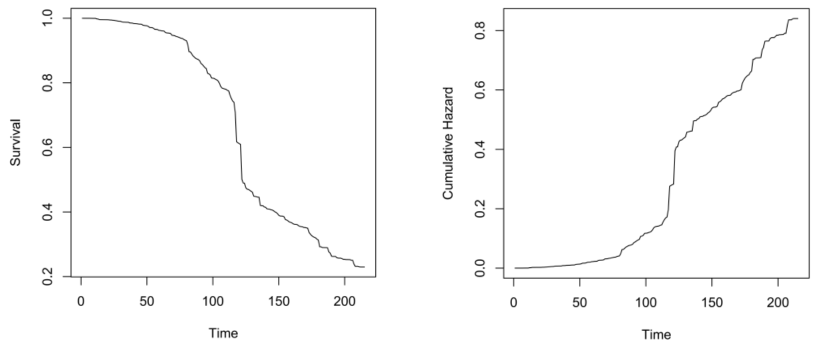
The patent CN105549647B, which discloses a mobile piglet traction local culture environment monitoring system, was applied for on 15 December 2015 by Zhejiang University and transferred to Hefei Shenmu Information Technology Co., Ltd. (Hefei, China) on 3 December 2021, with a lifetime of about 72 months. Put the individual patent data into the trained random survival forest model, and calculate its survival function as well as cumulative risk function through its feature indicators, as shown in Figure 5. From the survival curve in Figure 5, it can be observed that the survival rate of the patent drops sharply after about 70 months after its application, which indicates that the probability of patent transfer increases acutely at this point in time. Additionally, the cumulative risk function of the patent also rises abruptly at around 70 months after its application, indicating that the probability of the patent being transferred during this period is quite high. This prediction result is consistent with the fact that the patent is transferred at about 72 months after application.



(**A**) (**B**)

Figure 5. Survival function (**A**) and cumulative risk function (**B**) of patent CN105549647B.

Patent CN100386728C is a planned type of medical treatment instrument, which was applied on 24 March 2006 by Tsinghua University and transferred to Beijing Pinchi Medical Equipment Co., Ltd. (Beijing, China) on 26 October 2016, with a survival time of about 127 months. Put the individual patent data into the trained random survival forest model, and calculate its survival function as well as cumulative risk function through its feature indicators, as shown in Figure 6. It can be seen from Figure 6 that the probability of the patent being transferred at about 120 months after the application is less than 50%, which is due to the steep decline of the patent's survival curve and the steep rise of the cumulative risk function at about 120 months after the application, indicating that the probability of the patent being transferred during this period is fairly high, and this prediction outcome is in accord with the fact that the patent is transferred at sbout 127 months after its application.



(**A**) (**B**)

Figure 6. Survival function (**A**) and cumulative risk function (**B**) of patent CN100386728C.

To sum up, the survival function and cumulative risk function can be calculated according to the characteristic index of the patent, so as to judge the probability of the patent transfer at different time nodes.

6. Discussion

Clarifying the fluctuation of patent transfer probability over time can lead universities to seek potential patent grantees in the market at an appropriate time, and then realize patent industrialization effectively, facilitating the close integration of science and technology with economy. Considering that the existing relevant research was mainly focused on studying the factors affecting patent transfer cycle, less consideration has been given to the change of the probability of patent transfer over time. In addition, the dynamic period of patent transfer has not been predicted. Based on this background, time-to-event outcomes can provide more information than simply whether or not an event occurred. In order to deal with these outcomes, as well as censored observations where the event was not observed during follow-up, survival analysis methods are adopted in this research.

In an effort to select the model with the optimal prediction performance in university patent transfer cycle, Cox proportional risk model, Cox model based on lasso penalty, random forest model and random survival forest model are compared in predicting performance when applying the sample data of university patent transfer cycle. It shows that the prediction performance of the random survival forest is optimal among the four survival models, which can provide suggestions for both universities and enterprises on identifying the opportunity of patent transformation, thereby shortening the patent transfer cycle and improving the patent transfer efficiency.

However, there are also some limitations in this study: Firstly, patents in different technical fields may have different patterns in their transfer cycles; all kinds of university invention patent data are used for modeling as an integration in this paper, which can ensure the robustness of the model to some extent, but the prediction accuracy of an individual patent might be reduced. Secondly, this study simply takes the patent applied by C9 League for the model establishment as research samples, whereas the types and levels of different university patents are varied. Therefore, considering the finiteness of the sample types in this research, the scope of model application may have some limitations. In view of these deficiencies, further research is supposed to be performed in future work.

7. Conclusions and Suggestions

7.1. Conclusions

In this paper, the invention patent data of C9 League are taken as the research sample, and 12 specific variables are selected from the technical dimension, legal dimension and market dimension of the patents, respectively. Additionally, random survival forest model is introduced into the study of university patent transfer cycle and its performance is compared with that of the Cox proportional risk model, Cox model based on lasso penalty and random forest model. The conclusions are drawn as follows:

(1) The VIMP is calculated to obtain the importance ranking of each variable. It is concluded that the three indicators which affect the university patent transfer cycle most are the number of countries citing, non-patent citations and backward citations. The three indicators reflect the subsequent improvements of current patents made based on existing patents, and are essential for the evaluation of patent transfer cycle. The number of inventors, technical width, claims, family country, forward citation, countries cited and family size are indicators that also are related to the university patent transfer cycle, and have a certain effect on the prediction results of the model. Whether the patent is submitted through Patent Cooperation Treaty or whether it has experienced litigation events has little impact on its transfer cycle.

(2) The prediction result based on the test set of data illustrates that the prediction performance of the random survival forest model is superior to that of Cox proportional risk model, Cox model based on lasso penalty and random forest model by calculating and comparing the model evaluation indicators, which include C-index, Brier score and integrated Brier score. Moreover, the survival function and cumulative risk function that are generated through the random survival forest model provide the dynamic time-point prediction of an individual university patent transfer cycle, which indicates the validity of the random survival forest model.

7.2. Suggestions

Based on the conclusions above, the following suggestions are proposed to improve the transformation efficiency of university patent:

(1) It is necessary for universities to make subsequent improvements for current patents based on the existing patents, to strengthen the scientific research teams construction and enhance patent layout for core technologies.

Besides satisfying the modeling requirements, obtaining the variable importance of university patent transfer cycle indicators also provides reasonable guidance for patent application and maintenance. More specifically, university patent applicants need to increase the number of countries citing, non-patent citations and the backward citations to reinforce the technological innovation features for the patent. Moreover, the patent layout is supposed to be well conducted based on the core technology to form a certain technological advantage through patent applications from the same family, laying a foundation for increasing the patent forward citations afterwards. It is also essential for universities to expand the scale of scientific research teams to improve the knowledge and technology intersections of technical innovation personnel, thus enhancing the technology complexity and market competitiveness of the patents.

(2) Universities are supposed to strengthen the contact with enterprises to promote the scientific and technological cooperation between them, seizing the advantageous opportunity to promote patent transformation.

Random survival forest model can help universities and enterprises to identify the advantageous opportunity of patent transfer effectively. According to the survival function and cumulative risk function predicted by random survival forest model, when the probability of patent transfer is high, it is necessary for universities to increase exchanges with enterprises to promote patent industrialization by seeking potential patent grantees in the market through the patent technology transfer platform, technology product exchange meetings and intermediary institutions for science and technology transformation. Meanwhile, enterprises can introduce patent technology at an appropriate time as well, thereby achieving technology preoccupation prior to their competitors and improving the market competitiveness of their products and services. Furthermore, patents with low transfer probability and long period can also be identified by the random survival forest model and could be considered to be abandoned in time to reduce the patent maintenance costs.

Author Contributions: Writing—original draft, D.D.; Writing—review & editing, T.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Hubei Provincial University Teaching Reform Research Project: Culture Construction and Practice of Science and Engineering Universities from the Perspective of Higher Education Quality Assurance (no. 2018243) and the Key Project of Hubei Industrial Policy and Management Research Center: Research on the Service System of Industrial Technology Transfer in Hubei Local Universities (no. 17CYZ01).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Liu, J.; Zhong, Y.H. Research on the technology transfer based on patent licensing of the enterprises listed on the sci-tech innovation board. *Stud. Sci. Sci.* **2021**, *39*, 892–899.
2. Cheng, L.L.; Jiang, H.; Jin, J.; Xu, L.; Wang, X.Y.; Zhou, Z.H.; Wang, Y.; Zhu, Z.Z.; Zhou, J. The investigation of hospital patent transformation using multiple transactions. *Chin. J. Med. Sci. Res. Manag.* **2015**, *28*, 323–325, 328.
3. Jun, S.; Uhm, D. A Predictive Model for Patent Registration Time Using Survival Analysis. *Appl. Math. Inf. Sci.* **2013**, *7*, 1819–1823. [CrossRef]
4. Dutt, R.; Rathi, P.; Krishna, V. Novel mixed-encoding for forecasting patent grant duration. *World Pat. Inf.* **2021**, *64*, 102007. [CrossRef]
5. Sun, Z. Paradox of Chinese individual patent maintenance—Evidence of frictions in the patent market. *Sci. Res. Manag.* **2019**, *40*, 179–187.
6. Wang, Y.D.; Zhu, R.J.; Qi, Y.R.; Jin, J.; Huang, J.S. A state on the status and characteristics of external technology selection of Chinese firms. *Sci. Sci. Manag. Sci. Tech.* **2014**, *35*, 77–84.
7. Lee, S.S.; Lim, S. An Analysis on the Relation between Patent Maintenance Period and Patent Vale, and its Implication. *J. Intellect. Prop.* **2013**, *8*, 39–56. [CrossRef]
8. Trappey, A.; Trappey, C.V.; Wu, C.Y.; Lin, C.W. A patent quality analysis for innovative technology and product development. *Adv. Eng. Inform.* **2012**, *26*, 26–34. [CrossRef]
9. Gambardella, A.; Giuri, P.; Luzzi, A. The market for patents in Europe. *Res. Policy* **2007**, *36*, 1163–1183. [CrossRef]
10. Muscio, A. What drives the university use of technology transfer offices? Evidence from Italy. *J. Technol. Transf.* **2010**, *35*, 181–202. [CrossRef]
11. Fong, P.S.W.; Chang, X.; Chen, Q. Faculty patent assignment in the Chinese mainland: Evidence from the top 35 patent application universities. *J. Technol. Transf.* **2015**, *43*, 69–95. [CrossRef]
12. Llor, A. Delay from patent filing to technology transfer: A statistical study at a major public research organization. *Technovation* **2007**, *27*, 446–460. [CrossRef]
13. Mccarthy, I.P.; Ruckman, K. Licensing Speed: Its Determinants and Payoffs. *J. Eng. Technol. Manag.* **2017**, *46*, 52–66. [CrossRef]
14. Danish, M.S.; Ranjan, P.; Sharma, R. Determinants of patent survival in emerging economies: Evidence from residential patents in India. *J. Public Aff.* **2020**, *21*, e2211. [CrossRef]
15. Hemant, I.; Udaya, B.K.; Eugene, H.B.; Michael, S.L. Random survival forests. *Ann. Appl. Stat.* **2008**, *2*, 841–860.
16. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
17. Sun, J. *The Statistical Analysis of Interval-Censored Failure Time Data*; Springer Science + Business Media Inc.: Berlin/Heidelberg, Germany, 2006.
18. Lu, W.; Pan, X.; Dai, S.; Fu, D.; Hwang, M.; Zhu, Y.; Zhang, L.; Wei, J.; Kong, X.; Li, J.; et al. Identifying Stage II Colorectal Cancer Recurrence Associated Genes by Microarray Meta-Analysis and Building Predictive Models with Machine Learning Algorithms. *J. Oncol.* **2021**, *2021*, 6060762. [CrossRef]
19. Baralou, V.; Kalpourtzi, N.; Touloumi, G. Individual risk prediction: Comparing random forests with Cox proportional-hazards model by a simulation study. *Biom. J.* **2022**. *online ahead of print*. [CrossRef]
20. Chang, E.; Joel, M.Z.; Chang, H.Y.; Du, J.; Khanna, O.; Omuro, A.; Chiang, V.; Aneja, S. Comparison of radiomic feature aggregation methods for patients with multiple tumors. *Sci. Rep.* **2021**, *11*, 9758. [CrossRef]
21. Jeffery, A.D.; Dietrich, M.S.; Fabbri, D.; Kennedy, B.; Novak, L.L.; Coco, J.; Mion, L.C. Advancing In-Hospital Clinical Deterioration Prediction Models. *Am. J. Crit. Care* **2018**, *27*, 381–391. [CrossRef]
22. Qi, J.; Yan, J.; Idrees, M.; Almutairi, S.M.; Rasheed, R.A.; Hussein, U.A.; Abdel-Maksoud, M.A.; Wang, R.; Huang, J.; Huang, C.; et al. Five EMT-Related Gene Signatures Predict Acute Myeloid Leukemia Patient Outcome. *Dis. Markers* **2022**, *2022*, 7826393. [CrossRef] [PubMed]

23. Singh, R.S.; Totawattage, D.P. The Statistical Analysis of Interval-Censored Failure Time Data with Applications. *Open J. Stat.* **2013**, *3*, 155–166. [CrossRef]
24. Adhikary, A.C.; Rahman, M.S. Firth's penalized method in Cox proportional hazard framework for developing predictive models for sparse or heavily censored survival data. *J. Stat. Comput. Simul.* **2021**, *91*, 445–463. [CrossRef]
25. Gao, Z.; Ke, A.X.; Yu, R.J.; Wang, X.S. Urban Expressway Traffic Incident Duration Prediction Based on Random Survival Forests. *J. Tongji Univ. Nat. Sci.* **2017**, *45*, 1304–1310.
26. Wang, H.; Li, G. A Selective Review on Random Survival Forests for High Dimensional Data. *Quant. Bio-Sci.* **2017**, *36*, 85–96. [CrossRef] [PubMed]
27. Fang, K.N.; Zhang, G.J.; Zhang, H.Y. Individual Credit Risk Prediction Method: Application of a Lasso-logistic Model. *J. Quant. Tech. Econ.* **2014**, *31*, 125–136.
28. Adeoye, J.; Hui, L.; Koohi-Moghadam, M.; Tan, J.Y.; Choi, S.-W.; Thomson, P. Comparison of time-to-event machine learning models in predicting oral cavity cancer prognosis. *Int. J. Med. Inform.* **2021**, *157*, 104635. [CrossRef] [PubMed]
29. Chen, H.L.; Yu, S.J.; Xu, Y.; Yu, S.Q.; Zhang, J.Q.; Zhao, J.Y.; Zhu, B. Artificial Neural Network: A Method for Prediction of Surgery-Related Pressure Injury in Cardiovascular Surgical Patients. *J. Wound Ostomy Cont. Nurs.* **2018**, *45*, 26–30. [CrossRef]
30. Schneider, C.; Aubert, C.E.; Del Giovane, C.; Donzé, J.D.; Gastens, V.; Bauer, D.C.; Blum, M.R.; Dalleur, O.; Henrard, S.; Knol, W.; et al. Comparison of 6 Mortality Risk Scores for Prediction of 1-Year Mortality Risk in Older Adults with Multimorbidity. *JAMA Netw. Open* **2022**, *5*, e2223911. [CrossRef]
31. Morillas, N.R. La calidad del sistema universitario en China: La Liga C9 y la clasificación de Shanghái. *Estud. Asia Áfr.* **2019**, *54*, 203–228. [CrossRef]
32. Wei, T.; Liu, T. Evolution of High-Value Patents in Reverse Innovation: Focus on Chinese Local Enterprises. *Math. Probl. Eng.* **2020**, *2020*, 8127096. [CrossRef]
33. Xu, D.F.; Yu, H.D. How to operate patent: Innovation, market and law. *J. Chongqing Univ. Soc. Sci. Ed.* **2016**, *22*, 139.
34. Zhai, Y.; Chen, Y.; Teng, Z.; Zhao, Y. Identifying Antioxidant Proteins by Using Amino Acid Composition and Protein-Protein Interactions. *Front. Cell Dev. Biol.* **2020**, *8*, 591487. [CrossRef] [PubMed]
35. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
36. Xiong, J.; Shang, J. A penalized approach to mixed model selection via cross-validation. *Commun. Stat. Theory Methods* **2021**, *50*, 2481–2507. [CrossRef]
37. Reka, A.K.; Chen, G.; Jones, R.C.; Amunugama, R.; Kim, S.; Karnovsky, A.; Standiford, T.J.; Etherton-Beer, C.; Omenn, G.S.; Keshamouni, V.G. Epithelial–mesenchymal transition-associated secretory phenotype predicts survival in lung cancer patients. *Carcinogenesis* **2014**, *35*, 1292–1300. [CrossRef]