Computer-aided Analysis of Patents for Product Technology Maturity Forecasting

Yanhong Liang[1,2], Dequan Gan[1] , Yingchun Guo[2], Peng Zhang[1]

[1] Institute of Design for Innovation, [2] School of Computer Science and Engineering
Hebei University of Technology,Tianjin, 300130, P.R. China
{Yanhong.Liang, Dequan.Gan, Yingchun.Guo, Peng.Zhang, liangyanhong}@scse.hebut.edu.cn

Abstract. Product technology maturity foresting is vital for any enterprises to hold the chance for innovation and keep competitive for a long term. The Theory of Invention Problem Solving (TRIZ) is acknowledged both as a systematic methodology for innovation and a powerful tool for technology forecasting. Based on TRIZ, the state -of-the-art on the technology maturity of product and the limits of application are discussed. With the application of text mining and patent analysis technologies, this paper proposes a computer-aided approach for product technology maturity forecasting. It can overcome the shortcomings of the current methods.

Keywords: Product technology maturity foresting; TRIZ; text mining; patent analysis

1 Introduction

The analysis of the potential of emerging technologies and their implications, as well as the identification of the evolutionary stage of a product is vital for today's economies, societies, and companies. Such analysis affects critical choices made by multinational companies and individuals [1]. It is one of the most powerful guides to help determine the future development opportunities and limits for a wide variety of technical and business systems.

According to the results of Altshuller's studies, every product is a technical system supported by the core technology. The four typical stages of evolutional of a technical system are infancy, rapid growth, maturity and obsolescence, which make up of the technical lifecycle of product. TRIZ researchers have presented four primary indicators which can help to estimate the position of the system on the S-Curves [2]. The stage where a certain product locates in the S-curve is the product technology maturity. To determine the location of the product using S-curve is a critical activity in TRIZ evolution theory. The result of the technology maturity of a product can point out the decision-making direction for the enterprise. When a product is at infancy or rapid growth stage, its structures and parameters should be optimized to enter the maturity stage as soon as possible, in order to bring more profit to the enterprise. When a product is at maturity of obsolescence stage, it is still useful in the marketplace, still implemented, but one kind of substituted leading technology is

readily available and new product will be developed, which will make the enterprise gain competitive in the future.

Currently most methods of technology maturity forecasting rely on qualitative insights, and even if some quantitative methods emerge, such as statistical analyses (e.g., patent bibliometrics), they need participation of experts to a large degree. Actually, both qualitative and quatitative approaches have some shortcoming: they are cumbersome, difficult to apply, unreliable etc. In order to reinforce the weakness of one method it is combined with another method. However, when several methods are combined, new forecast errors can appear as a result of the synergy effects [3].

Patent documents contain important technical and commercial knowledge in terms of technical progress, market trend, and product innovation. If carefully analyzed, they can show technological details and relations, reveal business trends, inspire novel industrial solutions, or help make investment policy [4], [5].

In this paper, a computer-aided approach for identifying the product technology maturity is proposed. Based on modern information processing technology such as natural language process, statistical analysis and text mining, the aim of this paper is to find the intrinsic and objective laws from patents automatically and a quantified method of product technological maturity forecasting is established to overcome the shortcomings of the existing method of technology forecasting.

2   Review of Product Technology Maturity Forecasting

Based on his analysis of 2 million of the world's most successful patents, Altshuller recognized that the evolution of all technical systems is related to the performance characteristics of products, the level of patents, the number of patents and their profitability of products. Later it is applied to predict product technology maturity [6], [7]. However, the performance and profitability of products and level of patents are hard to access, which limit the application of this criteria method.

According to the basic functions of patents, Mann focused on examining two special types of patents - cost reduction patent and defect remedy patent [8]. By analysis of distributions of these patents in the product technical lifecycle, whether the product technology exceeded the maturity stage can be judged. But this method is obviously insufficient.

CHI's research group focused on studying the number of patents citations [9]. Patent forward and backward citations are commonly used to evaluate intellectual property competitive advantage. Besides, it normally takes five or more years from publication before a patent begins to be cited to any great extent. In general, 70% of all patents are either never cited, or cited only once or twice. So citation analysis is not suitable to technical fields with few ever cited patents.

From degree of the product to meet customer's demands Norman pointed out if the performance of product can satisfy customer's average demands the product is at maturity stage [10]. However, with the development of technology and the improvement of product performance, customer's demands are also changed, therefore the standard of technology maturity is necessary to change.

By integrating the research of Altshuller and Darrell Mann, Zhang huangao put forward a technique for product technology maturity mapping based on patent analysis [11]. The number of patents, level of patents, number of defect remedy patents are indictors for technology forecasting. Meanwhile he divided technical lifecycle to seven stages. The model of product technology maturity was built, and a software named TMMS was developed. The indicators relatively easy to access from the previous patents are left over, and the difficult indicators are abandoned. However, getting the level of patent and the patent for defect remedy is still dependent on participation of the experts, the model should be improved.

Cascini presented a computer-aided approach for technology forecasting by identifying the technical conflict overcome solved in a patent document on the basis of text mining technology [12].

There are some other methods for patent technology maturity forecasting. Some adopted the anagogic method of the number of patents to predict technology maturity. They determine the technology at maturity or rapid growth stage with technology maturity coefficient and rate. This method attempts to show technology maturity by the activity degree of innovation and the response to innovation of the enterprise, although it is simple but not accurate. Some experts predicted evolution by time series and the growth curve fitting, however the performance of product can't be expressed exactly, and accuracy of the prediction is under great influence on the selected mathematical model.

3   Technology Maturity Forecasting based on Patent-Mining

3.1   State of the Art of Patent-Mining Technologies

Product technology maturity forecasting is to study technology evolution, and technology innovation is the driving force of evolution, so the result of technology innovation contains information of technology evolution. Patent is the most common outcome of technological innovation, and patent document is a complete record of the patent activities, which can reflect the actualities of various technology activities and also can be used to study the evolution of the technology in a special technology field. At present time the number of worldwide patents has exceeded 5,400,000. The annual growth of patent application is about 4.75 percent in last decade. Patent documents contain important research results that are valuable for the process of product innovation.

Companies gain a strategic advantage over their competitors by watching the technology activities. Nevertheless, patent databases are large and complex and cannot simply be read. In fact, identifying the relevant inventions about the evolution of a specific technical system in a patent database is an indispensable task, since standard keyword-based search methods provide long lists of noisy results that require a careful manual selection. Thus, it is not easy, even for an experienced researcher, to recognize the importance of a patent and its relationship with other patents, especially when the corpus consists of hundreds of documents. With the development of

computational linguistics more and more text mining tools can provide an effective support to accomplish this task.

Text mining, which is used for drawing valuable information from large volumes of unstructured text, has been widely adopted to explore the complex relationship among patent documents. Text mining or knowledge discovery from text (KDT) - for the first time mentioned in Feldman et al. [13], is an emerging technology for analyzing large collections of unstructured documents for the purposes of extracting interesting and non - trivial patterns or knowledge. Text mining deals with the machine supported analysis of text. It has been proven as a means of uncovering hidden information in text-bases (electronic textual repositories). It uses techniques from information retrieval, information extraction as well as natural language processing (NLP) and connects them with the algorithms and methods of KDD, data mining, machine learning and statistics. Thus, one selects a similar procedure as with the KDD process, whereby not data in general, but text documents are in focus of the analysis.

Among all the techniques for technology forecasting, a growing acknowledgement is given to the TRIZ-based approaches. The analysis is based on the number of the patents, their innovation levels, their performance and profitability. Nevertheless, identifying the stage of evolution of a given technical system according to the criteria and all derivative method, is a rather complex and time-consuming task. While number-of-patents versus time functions can be easily, automatic innovation level and performance evaluations are very challenging.

Based on this, four indicators related to patents are selected for technology maturity forecasting in this paper.

3.2 Indicator Selection

1. The number of patents: It reflects the activity degree of technology research; its S-Curve is shown as Fig.1.a.
2. The frequency of patents related to keywords in documents: The performance of product as one of the indicators for product technology maturity forecasting, suggested by Altshuller, is logical, but the criteria for practical usage is impossible to provide and therefore is extremely difficult to quantify. Therefore, it is necessary to find an alternative indicator.

The keywords of a document can reflect its topic. So the study hot spots of the specific field can be discovered by analysis of keywords. That is to say, a domain word that appears in a document with high frequency indicates that the word is a significant keyword, which is the basis of the indicator proposed in this paper. It denotes the higher frequency of patents related to the keywords for a period of time, the technology related to these words is more active.

As there are a great deal of words in the patent collection, each word should be summarized. The higher frequency of words will be selected. And then each word will be taken as a node, the frequency of two words appear at the same time will be calculated and the higher frequency of the correlative words are taken as keywords. Here the associating rule can be used to extract the key words. The number of patents including these keywords is named as frequency. Frequency of the patents versus time

series can represent the research hotspot of the specific field. Its S-curve is similar to the performance in general, as shown in Fig.1.b.

3. The degree of patent novelty: It is first mentioned for technology forecasting by Liu Yuqin [14]. Novelty is usually referred to one of the most important criteria that an applicant must demonstrate for an invention before a patent can be granted. It is a patentability requirement. An invention is not patentable if the claimed subject matter was disclosed before the date of filing, or before the date of priority if a priority is claimed, of the patent application. Since in the context of subsequent innovations each innovation is built on the previous, the degree of novelty is an indicator to evaluate the level of patent, which is based on document similarity in this paper.

Patent documents are splited into bag-of-words automatically. There are a number of approaches to extract words relevant to the same topics from the patent collection. For given a collection of n patent documents, each patent document can be denoted to Vector Space Model (VSM) form, such as: $D_j = (td_{i1}, td_{i2}, \cdots td_{in})$ .Here td_{ij} is weight of the word i in the patent j. Similarities are then computed among all useful word pairs. A typical similarity measure is given by cosine function (1):

$$Sim(D_j, D_k) = \sum_{i=1}^{n} td_{ij} \times tq_{ik} \left/ \sqrt{\sum_{i=1}^{n} td_{ij}^2 \sum_{i=1}^{n} tq_{ik}^2} \right. \qquad (1)$$

If the weights of the words were either 1 to 0, denoting the presence or absence of the words in patent documents, the similarity becomes a value exactly proportional to the number of documents in which these two words co-occur. The weight of word can be computed automatically on according to term frequencies (TF), or TF $\times$ Term_Length, or other criterions such as TF$\times$IDF (Inverse Document Frequency). The equation of TF-IDF is shown as in (2):

$$TF - IDF(t_j, d_i) = (tf_{ij} / idf_j) \bullet (\log_2(n/n_j) + 1) \qquad (2)$$

Here, tf_{ij} is the number of word t_j which occurs in patent i. idf_j is the biggest number of some word which occurs in patent i. n is size of the total patents collection. n_j is the number of the patents collection that contain word t_j .

The degree of patent novelty is given by function (3):

$$Nov_\alpha(P) = e^{-\frac{n}{2}} \qquad (3)$$

Here, n is the number of patents which similarity is bigger than α before the time P .

The number of patents similar to a patent is more, the worse the novelty of the patent is. The lower the value of novelty is, the higher the degree of novelty is. The degree of novelty is similar to patent citation, which can be used to evaluate intellectual property competitive advantage. Patent citation is to inspect a patent referenced by the followers, while the degree of novelty is to inspect a patent

influenced by the prior patents. But 70% of all patents are either never cited, or cited only once or twice, which limit the application of patent citation. The degree of novelty overcomes this defect, which is used to search potential rule from the text. The degree of novelty indicates the quality level of patent, and the shape of its S-Curve is analogous to that of the level of patent versus time. It is shown as Fig.1.c.

4. The maintenance cost of patent: Altshuller took the profitability of products as one of the indicators for product technology maturity forecasting. Considering economic interest of the enterprise, it has been greatly hindered in the practical application. Therefore, an alternative indicator is proposed.

Schankerman and Ariel put forward the condition of patent maintenance - the return should at least equal to the maintenance cost of patent each year [15]. In most countries patentees must pay an annual renewal fee in order to keep their patents in force. This information can be accessed by the data of patent legal status. Therefore, the change of patent maintenance cost reflects the change of product profitability. The maintenance cost of patent can be taken as one of indicators for technology maturity forecasting. Its S-curve is similar to the profitability in general, as shown in Fig.1.d. The initial cost of patent is not a negative, since patentees must pay application fee when they apply for a patent, which is different between them.
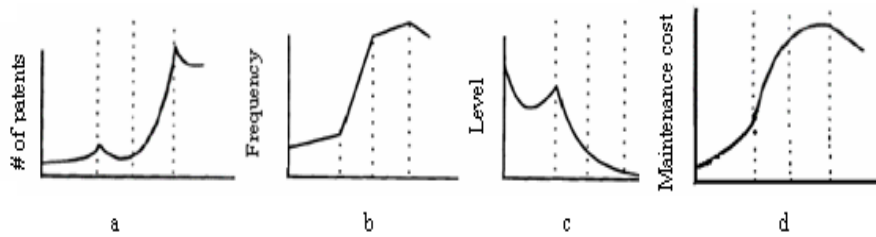

Fig. 1. S-Curve of technology maturity forecasting

3.3   Determination of Product Technology Maturity

Each indicator as mentioned above has its own characteristic S-Curve (in time for the abscissa). In Fig.1, the abscissa is divided into four stages - infancy, growing, maturation with three dotted lines. From the map we can see system changes in four-scale variables at the four stages. The product technology maturity can be determined by the combination of four indicators under the history observation points of patent collection.

In practice, four variable curves in the specified field are depicted with the collection of relevant data, and they will be compared with the characteristic curves of the scale variables, then the stage of the technology evolution can be determined.

4   Process of Product Technology Maturity Forecasting based on Patent-Mining

On the basis of patent analysis and text mining technologies for technology forecasting, the patent documents should be pre-processed and statistical analyzed. The overall process of conducting product technology maturity forecasting goes through several steps.

First of all, data collection and data preprocessing are the preliminary step. The patent area of interest is selected and related patent documents are collected in electronic text format. Many patent offices already allow to downloading abstracts and completed text of the patents freely. We can search the patents under some IPC categories or keywords. This may involve a repeated process of devising a set of query terms, searching a couple of patent databases, filtering undesired patents, and downloading patents to local area. And in order to get the newest legal status of patents, these patent collections will be searched under legal status.

Second, on the basis of patent abstracts or claims, raw patent documents are transformed into structured data. Since the original documents are expressed in natural language format, they must be transformed into structured data in order to be analyzed and utilized. Text mining that extracts keywords from patent document is used to this end.

The sentences of the patent document are split into words automatically and stop words are filtered with the aid of computer. There are some undesired words such as function words which will impact the final output and analysis. To better match meanings of words, words are stemmed on Porter's algorithm [16]. These words are then used to construct a vector space model. In this step, the document is represented by a vector containing the number of selected words. Words in each patent document are first sorted in decreasing order of their term frequencies or TF$\times$IDF. Then the first k words are selected for analysis. The patent similarity is calculated, and after the level of similarity is given, the patent novelty is calculated.

Third, the keywords are extracted from patents using the associating rule of text mining. A word correlations matrix is constructed for synthesizing two or more highly co-related words into a smaller set of significant words. The frequency of patents related to these keywords is calculated.

The second and third steps are pivotal for technology forecasting. The forecasting result is direct affected by the quality of participle, stop words considered, method of weights of words for VSM, the selected algorithm and so on.

Fourth, the annual maintenance cost of patents is calculated according to the patent legal status.

Fifth, these patent data mentioned above are collected according to a certain period of time. The number, frequency, novelty and maintain cost of patents are summed in unit time. Four curves on the basis of indicators versus time in the specified field are depicted with the collection of relevant data. Then they will be compared with the standard curves, the S-Curve position of the technology can be determined.

Finally, results will be returned to the relevant technical experts in the specific field. The experts will evaluate the forecast results with the comprehensive of the current product technology performance, profitability and other indicators.

5   Conclusion

The technology maturity of product forecasting is assuming a crucial role for improving the competitiveness of any companies. Therefore the development of methodologies and tools for assessment of the evolutionary stage of a give product automatically, or at least computer-aided analyses of patents is in a growing demand. The state-of-the-art on technology forecasting and the limit of appliance were discussed. The TRIZ approach to technology forecasting is gaining acknowledgement, whereas the human-based assessment is rather complex and time consuming. In order to overcome the limitation, in this paper a set of quantitative indicators of weights to substitute the traditional measures and a method to estimating a system technology maturing is proposed. The approach for technology maturity of product forecasting is integrated on patent analysis and text-mining techniques.

Now a test is conducting to verify this method. And further efforts will be dedicated to optimize the algorithms for VSM and keywords extraction to improve the prediction accuracy.

Acknowledgement

References

1. Alan, L.P.: Technology futures analysis: Toward Integration of the Field and New Methods. Technological Forecasting & Social Change. no. 71, pp. 287–303 (2004)
2. Runhua Tan: Theory of Inventive Problem Solving. Science Press (in Chinese). Beijing (2004)
3. Dmitry, K., Roland, D.G.: Problems of Forecast. Proceeding of the ETRIA TRIZ Future 2005.Graz, Austria, 16-18 November, pp. 219-233 (2005)
4. Campbell, R.S.: Patent Trends as a Technological Forecasting Tool. World Patent Information. 5(3), pp. 137－143 (1983).
5. Jung, S.: Importance of using patent information. In WIPO—Most intermediate training course on practical intellectual property issues in business, organized by the World Intellectual Property Organization (WIPO), Geneva, November 10－14.) (2003)
6. Frauens, M.W.: Improved Selection of Technically Attractive Project's Using Knowledge Management and Net Interactive Tools. Cambridge：MIT (2000)
7. Slocum, M.S.: Technology Maturity Using S–curve Descriptors. http://www.triz-journal.com/archives/1998/12/a/index.htm (2008)
8. Darrel,l M.: Using S-Curves and Trends of Evolution in R&D Strategy Planning.

http://www.triz-journal.com/archives/1999/07/g/ (2008)

9.NARIN,   F.:   Tech-Line   Background   Paper.
http://citeseer.ist.psu.edu/cache/papers/cs/24773/http:zSzzSzchiresearch.comzSztechlinezSzt
lbp.pdf/unknown.pdf (2008)

10. Norman, D.: The Life Cycle of a Technology: Why It is So Difficult for Large Companies
to Innovate..http://www.jnd.org/dn.pubs.html (2008)

11. Huagao Zhang, Weiyan Zhao, Runhua. Tan: Technique of Product Technology Maturity
Mapping based on Patent Analysis and Software Design (in Chinese). China Mechanical
Engineering. vol.17, no. 8, pp.823-827 (2006)

12. Cascini, G., Russo, D.: Computer-aided Analysis of Patents and Search for TRIZ
Contradictions. International Journal of Product Development. vol. 4, no.1/2, pp. 52-67
(2007)

13. Feldman, R., Dagan, I.: Kdt - knowledge Discovery in Texts. In Proc. of the First Int. Conf.
on Knowledge Discovery (KDD). pp. 112–117 (1995)

14. Yuqin Liu, Donghua Zhu, Lin Lv.: Technology Maturity of Product Forecasting based on
Text Mining (in Chinese). Computer Integrated Manufacturing Systems. vol.14, no.3,
pp.506-510 plus pp.542 (2008)

15. Schankerman, M., Pakes, A.: Estimates of the Value of Patent Rights in European Countries
During the Post-1950 Period. The Economic Journal (1986)

16. Porter, M. F.: An Algorithm for Suffix Stripping. Program. vol.14, no.3, pp.130-137 (1980)