

© 2013 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

A Time-series-based Technology Intelligence Framework by Trend Prediction Functionality

Hongshu Chen ^{a,b}, Guangquan Zhang ^a, Jie Lu ^a

Decision Systems & e-Service Intelligence Lab ^a

Centre for Quantum Computation & Intelligent Systems

Faculty of Engineering and Information Technology, University of Technology, Sydney

PO Box 123, Broadway NSW 2007, Sydney, Australia

School of Management and Economics ^b

Beijing Institute of Technology, Beijing, China

hongshu.chen@student.uts.edu.au, {guangquan.zhang, jie.lu}@uts.edu.au

Abstract—Technology Intelligence (TI) indicates the concept and applications that transform data hidden in patents or scientific literature into technical insight for technology development planning and strategies formulation. Although much effort has been put into technology trend analysis in existing research, the majority of the results are still obtained from expert opinions on the basis of historical trends presented by content-based Technology Intelligence tools. To improve this situation, this paper proposes a time-series-based framework for TI that enables the system to be more effective when dealing with trend prediction requirements. Time-series analysis module is first applied in TI framework to process patent time series for technology trend predictions in a real sense, at the same time overcome the problem that prediction of future data points' values is insufficient to support TI construction. Based on explicit patent attributes and unknown patterns learned from the historical data, the framework combines the “trend” and “content” knowledge by analyzing both time-related property and semantic attributes of patent data, to support technology development planning more efficiently and satisfactorily. A case study is presented to demonstrate the validity of trend prediction functionality, which is the emphasis of the whole framework.

Keywords—Technology Intelligence; patent analysis; technology trend prediction

I. INTRODUCTION

Modern society increasingly emphasizes the role of technology research and development (R&D), promoting technical advances and accumulation of intellectual property to a great extent. Aiming to obtain more opportunities, avoid potential risk as much as possible and track new innovation trends, the demand for efficient technology strategies is becoming increasingly important for future technological development in both public and private domains. Traditionally, experts play a principal role in technology strategy formulation [1]. The data used in the majority of expert-based approaches and qualitative approaches summarized as ‘soft’ methods to support technology analysis [2], come from surveys and questionnaires, which are based on the experience of experts who answer them and researchers who design them. Thus, the data used in these approaches is unavoidably influenced by subjective opinions of experts. At the same time, the collection procedures are

usually expensive and time consuming for users. However, strategy makers require increasing technical support for high quality decisions in a timely manner [3]. In order to improve the situation, the concept and tools of *technology intelligence* are proposed to improve the traditional expert-experiences dominant process by using objective data and artificial intelligence techniques for technology analysis.

The concept of *technology intelligence* (TI) was first systematically mentioned in supplier management research [4]. Just like Business Intelligence to support business decision making, TI is promising to turn ‘data’ found in patents or scientific literatures into ‘knowledge’, and help users survive data tsunamis and eventually, to succeed in strategy making [5]. Reviewing existing studies, TI can be viewed as an ‘activity’ to be conducted by a set of agents, or as a knowledge management ‘product’ with consumers, that provides an organization with the capability to capture and deliver information in order to develop an awareness of technology threats and opportunities [6]. Compared to the traditional expert-based approaches, TI utilizes objective data (patents or scientific literatures), not expert grading, as the fundamental data source, which makes the results more unbiased. In addition, it is capable of dealing with larger quantities of information that are not able to be analyzed by humans alone [7], and also has capability to generate knowledge by integrating resources from different sources to visualize the outcomes. TI concepts and applications make it possible to monitor the direction of technology growth and plan for technology R&D for both government and private companies [8].

Since this concept was first presented, research of TI gradually expanded from statistics-based to data-mining-based and became more intelligent in recent years. That is, an increasing number of researchers focused on use of extremely powerful information technologies and a vast amount of available data that digitally provides us with technology intelligence [9]. *TrendPerceptor* [7], is designed to identify TRIZ (Russian Acronym of the Theory of Inventive Problem Solving) trends in invention concepts by using a property-function based approach. *Techpioneer* [8] uses text mining and morphology analysis to seek potential technology opportunities. *VantagePoint* and *Aureka* are another two intelligent systems that support users to analyze

trends or relationships by providing clustering, mapping and searching techniques [10, 11]. There are also some applications built based on bibliometric approaches, for instance, a visualization system called *Diva* was proposed to perform bibliometric analysis of scientific literature and patents for trend presentation [12]. In summary, the existing design and tools of TI are mainly constructed on the fundamentals of semantic properties of patent documents and scientific literature. The trends/reasons/keynotes/changes of a particular technological area presented by these methods mainly focused on current and historical states. Although ‘technology forecasting’ was mentioned in many previous studies frequently, time-related prediction was seldom involved when a technology intelligence structure was considered.

As a remedy, this paper proposes a time-series-based TI framework with trend prediction functionality using patent data. Time-series analysis module is first applied to TI framework to process patent time series and overcome the problem that prediction of values in future time intervals is insufficient to support TI construction. In addition, semantic attributes and time-related attributes of patents are considered at the same time to present ‘content’ and ‘trend’ knowledge, correspondingly. Based on explicit patent properties and unknown patterns learned from the historical data, we combine the “trend” and “content” knowledge to support technology development planning and strategy more efficiently and satisfactorily.

After the introduction, this paper is organized as follows: Section 2 presents a new solution for TI by illustrating the appropriate input data source and output knowledge of our framework. At the same time, the concept of Time-series-based Technology Intelligence (time-series-based TI) is proposed in this Section. In Section 3, we present the outline of the framework, and then explain the whole system architecture and its components in detail. Section 4 provides a case study to demonstrate that the core component of the framework, trend prediction functionality, can perform forecasting more efficiently and reasonably than traditional future values prediction. Conclusion and future studies are given in the last Section.

II. TECHNOLOGY INTELLIGENCE SOLUTION

A. Patent Time Series Data

Patents are the ideal data source for TI study. They stand for the intellectual properties of organizations ranging from individual ones, like companies, to the multinational level, like governments and unions. Since patents are provided to the public by patent offices of many countries around the world, it is not difficult to obtain patent data for academic study and commercial business. The research of using patents as indicator for technology analysis began in the early 1980s, and since then, patent analysis has become a very useful tool that supports technology R&D planning, competition analyses, and analytic studies of how technologies emerge, mature and disappear [13]. For standardization reasons, we use the primary source of patent data, the United States Patent and Trademark Office

(USPTO) database [14], which has been employed in the majority of studies.

As mentioned above, patent content-related research has been involved in technology intelligence framework design for many years. There is no doubt that patent documents containing abstracts, claims and description of details is one of the most significant data sources of the TI system formation. However, patents have both time-related property and semantic attributes [15]. The time-related property of patent data is also important, yet easy to be ignored. Within a particular industry, the quantity of patents under a certain search statement, such as relevant USPC (United States Patent Classification), IPC (International Patent Classification), a number of keywords or their combination during a period of time, can be presented as one vector. If we set time here at uniform intervals, the vector we obtain from patent publication history could be seen as time-series data. Let $P = \{p_1, \dots, p_m\}$ defines one sequence, p_i ($1 < i < m$) indicates the patent quantity that appears in each corresponding time interval, while n stands for the number of intervals (seasons or years).

Users of TI are expected to perceive from patent documents not only what the content of the current and historical trend is, but also what further trends of these contents will perform in future. This requirement needs to be satisfied on the fundamental of time series analysis. Previous time-series-related research in the technology forecasting domain applied Linear Regression, Support Vector Regression (SVR), Moving Average (MA) model and Autoregressive Integrated Moving Average (ARIMA) model to conduct trend analysis [1, 16]. However, the above mentioned time series approaches were seldom considered when constructing a TI system because of two reasons:

- The exploration of future trend is more significant than the prediction of future data points’ values.

Most past studies focus on the future value prediction only. If we use patent time series directly to predict the future value, the result will be only the quantity for the next period of time, which is basically pointless for the TI systematic process.

- Patent publication/application activities are nonlinear.

As an indicator of technology development, patent publication and application behaviors could be seen as a nonlinear system with a number of nonlinear behaviors such as multi-stability and aperiodic oscillations. Thus linear time series prediction models have their limitation to perform accurate result.

Under this circumstance, a future trend or state, but not the exact number, is of more concern for TI framework construction. Aiming to forecast the future trend for one technology or several technologies, transformations between the count of publications and the states of development need to be learned from historical patent time-series data. In this research, we use Piecewise Linear Representation (PLR) [17] to decompose patent time series into a number of segments to

obtain the hidden trends and patterns in patents publication/application activities.

B. Time-series-based Technology Intelligence

We apply time series analysis to improve the existing concept of technology intelligence to develop the new one: Time-series-based TI. As shown in Fig.1, the existing concept of TI is built fundamentally of a combination, i.e. the intersection of artificial intelligence and technology strategy formulation. In order to satisfy the demands of finding more knowledge from patent documents, it then unites with patent content analysis, which helps to improve the system into a content-based TI level. The significance of our research is that we apply time series analysis to the framework, which provides the new system with the ability to process patent time-series data. To sum up, time-series-based TI is an intelligent system framework which utilizes time series analysis and patent content analysis to support the technology development planning and strategy process.

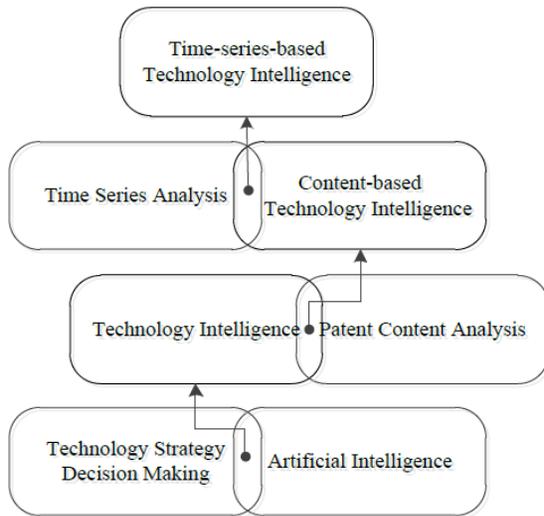


Figure 1. Concept of time-series-based Technology Intelligence

C. Outcome of the System

The outcome of the system consists of two sections, which derive from text mining and time series analysis, correspondingly.

For the first section, like previous content-based TI frameworks, we utilize text data that extracts from patent documents to present keywords clusters and relations hidden in massive patents of the target technology. For the second section, to improve the design of the content-based system, time series analysis is applied to the new framework. In order to forecast the future trend, but not the future number of patent publications, the result is derived from time series analysis and will be presented as a number of different states. The system will then unite the two parts of outcomes into a knowledge base.

The combination of two types of result provide users with patent knowledge that indicates: (1. what the historical and future patent publishing trend is in the technical area of interest; 2. where are the states turning points for a group of patents and what the next state will be; 3. keywords cluster

and relations to each state; 4. keywords cluster and relations for the whole trend; 5. what content (keywords cluster and relations alteration) most contributes to state transformation. This outcome will help users to better understand the content development, along with the time, and provides a prospect of the future trend/state in their technical areas of interest.

III. TIME-SERIES-BASED TECHNOLOGY INTELLIGENCE FRAMEWORK

After illustrating the system input and outcome, this Section will first provide an outline of our framework, and then describe the detailed architecture and its specific components.

A. Outline of the Framework

Users of TI are technology R&D managers of companies and technology planning officers in government sectors. For future technical development of their organizations, they all need to assess external technological developments to determine how they can gain from technology changes, avoid potential risk, and plan their future R&D activities [18].

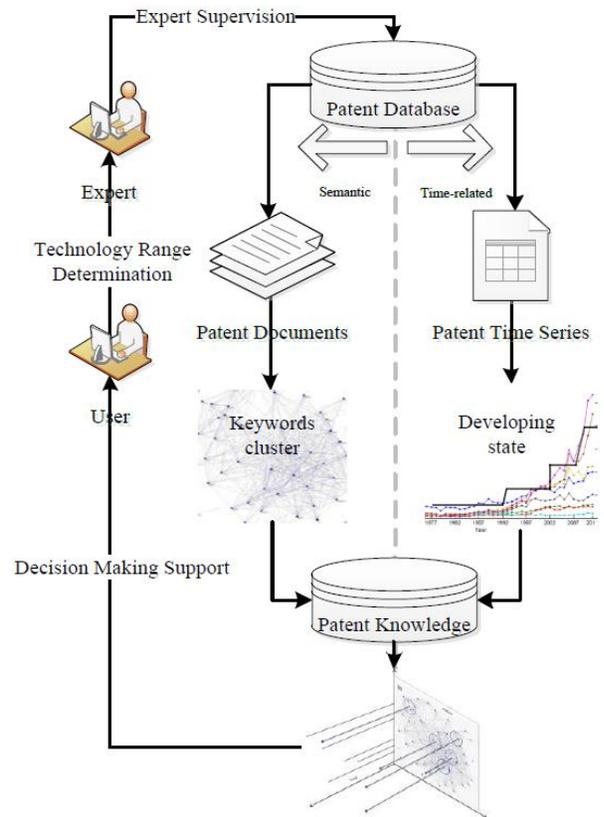


Figure 2. Conceptual model of time-series-based Technology Intelligence

The framework we proposed in this paper, i.e. the time-series-based TI framework, extracts patent time-series data and text data separately, then correspondingly processes them with trend forecasting and patent content analysis components. The whole framework builds on the fundamental of understanding, extracting and utilizing both time-related properties and semantic attributes of patent data, and learning patent knowledge from historical records. As

shown in Fig.2, users need to determine their technology area of interest as systems input at the beginning. The determination will then be turned into one or several patent query commands for the USPTO database under expert supervision; that is, we can obtain a group of patents that conform to the query functions such as IPC/USPC that relate to the target area, selected keywords that restrict the topic of the patents, and so forth. Although experts will still participate in system procedures, the effort this consumes confines them to supervise the input of users and provide appropriate advice about transformation between selected technology and patent search commands. After extracted from patent database, time series and text data will be computed and presented in keyword clusters and developing trends. The outcome of the system is patent knowledge that contains patent developing states and their corresponding keyword clusters, and also the prediction result that shows what state will be in the next time interval for the technologies of interest. The results will then be delivered to users for technology strategy support and future developmental planning.

B. System Architecture

Fig.3 shows the whole system architecture, which consists of two parts: *Technology Trend Forecasting* component and *Patent Keyword Clustering* component.

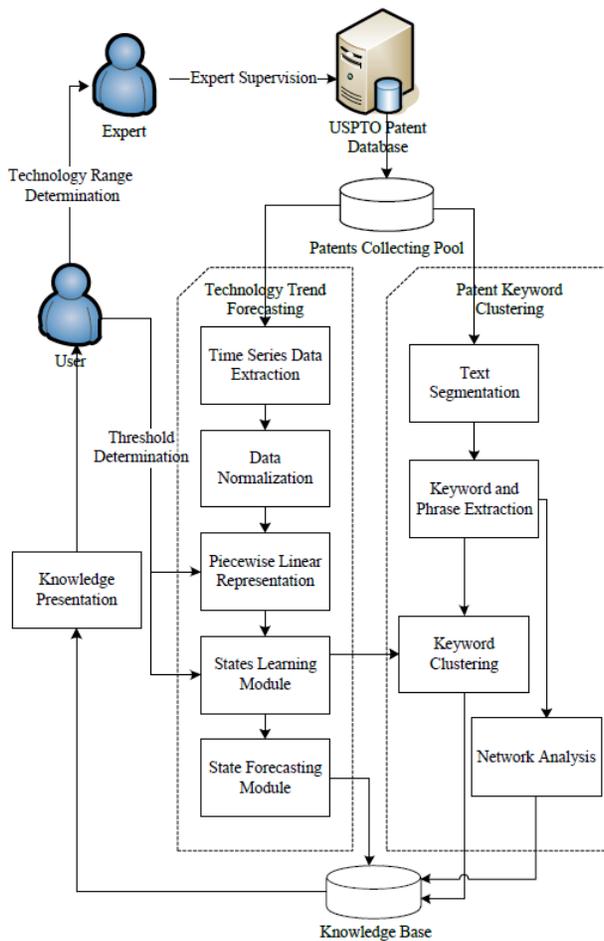


Figure 3. Architecture of time-series-based Technology Intelligence

Each component here is constructed of several modules with specific functionalities. After users of the system enter the patent query commands that define the technology range of their concern, all the patent documents that conform to the query statement will be searched and collected into a patent-collecting pool, and processed into patent time-series data and text data, separately. The component of Technology Trend Forecasting receives the time-series data. Because trend is of more concern here for users than single future value, the component will first simplify time-series data by using the Piecewise Linear Representation (PLR) approach, and then convert the outputs of PLR to possible states of patent time series. The historical states data thus can be used to predict new one for the next time interval. At the same time, the component of Patent Keyword Clustering receives the patent text data. By using text mining techniques, it clusters the keyword hidden in massive text data and figures out relationships between them. All result of the two components is logged into a knowledge base for further decision making support.

C. Technology Trend Forecasting Component

Technology trend forecasting is a significant part of technology intelligence. For technology intelligence users, it is important to compare the growth patterns of various technologies in a particular industry to help understand the industry characteristics and analyze the technology innovation process. At many points much effort has been put into technology trend analysis and presentation in existing research and study. However, most emphasize what the content of the current and historical trend is. When it comes to future prediction, the majority of the results are still obtained from expert opinion after the TI system presents current and historical trends to them.

In the time-series-based TI framework we propose in this paper, Technology Trend Forecasting component is the most significant part that showing the system characteristics. Time-related attributes of patents are first taken into consideration for accessing technology future trend prediction in a TI framework construction. As shown in Fig.3, selected patent data is placed in a data processing pool for the next-step analysis. Then the module of Time Series Data Extraction collects the time series and passes it to the Data Normalization module. After normalization, the time series dimension reduction tool, PLR module is used to simplify the normalized sequence and ready it for states learning. We already know that future technological impacts in a time period of interest can be directly assessed with use of future patent quantity counts [14]. The reason we utilize PLR here is that, patterns, not individual points, under our requirement of trend analysis are of much more interest. Users can give their preferred threshold for PLR, or it will be processed as the system default value, which maintains the balance of producing the smallest number of pieces and lowest RSS (residual sum of squares). After the PLR module, the trend of time-series data with a more obvious trend is placed in the States Learning module. A threshold (max angle of the slope to start a new state) is needed here to determine to what extent the trend changing shows it enters the next state from the current one. This study adopts a

transform function to transform the segments into trend states. Fig. 4 shows the pseudo code of the transform function. Here we set threshold, the angle of slope changing, as 30 degrees.

```

Input: the beginning and ending points of each segment
      P(bxn, byn), Q(exn, eyn)
Output: the states sequence S(n)
Normalize P(bxn, byn), Q(exn, eyn).
Calculate the angles for each segment as A(n).
P(bx1, by1) is the beginning of the state 1
  For i=1 to n
    If -30 < A(i) < 30
      Maintain the current state and record state in S(i)
      % the first slope is always positive
      i=i+1
    If A(i) ≥ 30
      P(bxi, byi) is set as the beginning of a new state n+1
      State+1 and record state in S(i)
      If A(i+1) < 0 and A(i+2) > 0
        Finish this state and i=i+1
        P(bxi, byi) is set as the beginning of a new state n+1
      Else
        Maintain the current state and Record state
    If A(i) ≤ -30 and A(i+1) > -30
      Finish this state and i=i+1
      P(bxi, byi) is set as the beginning of a new state n+1
    Else
      Maintain the current state and Record state
  End for

For each state
  Calculate the mean value of all the segments in it
  Merge the ones with smallest distance if need and refresh S
End for

```

Figure 4. The pseudo code for segments-states transformation

Once historical states are identified from input time series, the state data is processed to both the State Forecasting module and text analysis component. The former uses Neural Networks approach to predict the state in the following time interval. Finally, the result from the State Forecasting module will be placed in the knowledge base to support decision making of users.

D. Patent Keywords Clustering Component

Patent documents contain significant research results, and text-based analysis for assisting decision makers has been put into practice for many years [19]. In our proposed framework shown in Fig.3, a requirement initiated by the system users leads to a query in the USPTO patent database for a group of patents of a certain technological area. As mentioned above, the selected patent data will be placed in a data processing pool for further analysis. Then, the unstructured text data of patent documents is extracted and placed in the Text Segmentation module. The mission of this module is splitting the paragraphs and presenting text data in a matrix for the next-step clustering and relationship computing. However, there will be many more words than are actually required with different characteristics, thus the Keywords and Phrase Extraction module is utilized to identify those that are more important and meaningful for further study. In the following step, the system framework applies the Keywords Clustering module and Network Analysis module to identify and detect topics and relations in the keywords collection. Result from the States Learning module of Technology Trend Forecasting is also imported to this module, in order to compute the

keywords cluster and relations on each state and the whole trend. Moreover, by observing and comparing the clustering result of every state, we are able to obtain the knowledge about what content (keywords cluster and relations alteration) most contributes to state transformation. Finally, the outcomes will be input to a knowledge base.

IV. CASE STUDY AND EXPERIMENT

In this section we present a case study to demonstrate the validity of the core component of our framework, Technology Trend Forecasting, within real patent analysis context. The data we employ in the case study is created base on search statement for patents in USPC 455, which stands for Telecommunication technology. The time interval unit here is set as season. We choose the number of issued patents in every season during 1963 to 2012 to create time series, which makes 200 seasons in total.

In order to maintain the balance of producing the smallest number of pieces and lowest RSS, here we set the threshold δ as 20, which indicate the number of segments produced by bottom-up PLR algorithm. For every straight line shown in Fig.5, there are $(x_{t_{i-1}+1}, y_{t_{i-1}+1})$ and (x_{t_i}, y_{t_i}) indicate the beginning and ending point of the line during time t_i , where $1 < i \leq n$ and n stands for the total number of the segments. The slope is used to determine the possible states of the whole time series by using the segments-states transformation function as mentioned above. The normalized time series, the result present by PLR and the states presentation is shown as Fig.5. There are 5 states in all and the state turning years are 1990, 1998, 2005 and 2008.



Figure 5. The PLR result and states presentation for USPC 455 time series

Then we utilize the states of first 196 seasons to predict the final four, which stand for year of 2012. The experiment shows that the after transform the original time series into states, NNs performs very well in prediction. The forecasting result is shown as Fig.6. We already know that the trend of patent publication in USPC 455, Telecommunication technology area, has been rising quite fast since 2008, and this trend maintains in 2012. The empirical result for the state value of the 4 seasons in 2012 is 5.0043, which show the same trend as the actual situation. Furthermore, the high speed of the increase of telecommunication-related patents will continue in the next 4 seasons as well.

ACKNOWLEDGEMENT

The author Hongshu Chen acknowledges the support from the China Scholarship Council (2011603004) and top-up scholarship from University of Technology, Sydney.

REFERENCES

- [1] S. Jun, and D. Uhm, "Technology forecasting using frequency time series model: Bio-technology patent analysis," *Journal of Modern Mathematics and Statistics*, 2010. 4(3), pp. 101-104.
- [2] A. Porter, and P. Van Der Duin, "Technology futures analysis: toward integration of the field & new methods," *Technological Forecasting and Social Change*, Volume 71, Issue 3, March 2004, pp. 287-303
- [3] J. Lu, D. Ruan, and G. Zhang, "Guest Editorial: A special issue on Intelligent Decision Support and Warning Systems," *Knowledge-Based Systems*, 2010. 23(1), pp. 1-2.
- [4] R.D. Shapiro, *Toward effective supplier management: international comparisons*, Harvard Business School, 1985.
- [5] L.Niu, J. Lu, and G. Zhang, "Cognitive orientation in business intelligence systems," *Intelligent Decision and Policy Making Support Systems*, Springer Berlin Heidelberg, 2008, pp. 55-72.
- [6] C.I. Kerr, et al., "A conceptual model for technology intelligence," *International Journal of Technology Intelligence and Planning*, 2006. 2(1), pp. 73-93.
- [7] J.Yoon, and K. Kim, "TrendPerceptor: A property-function based technology intelligence system for identifying technology trends from patents," *Expert Systems with Applications*, 2012. 39(3), pp. 2927-2938.
- [8] B.Yoon, "On the development of a technology intelligence tool for identifying technology opportunity," *Expert Systems with Applications*, 2008. 35(1), pp. 124-135.
- [9] T.U. Daim, D.F. Kocaoglu, and T.R. Anderson, "Using technological intelligence for strategic decision making in high technology environments," *Technological Forecasting and Social Change*, 2011. 78(2), pp. 197-198.
- [10] D.Zhu, and A.L. Porter, "Automated extraction and visualization of information for technological intelligence and forecasting," *Technological Forecasting and Social Change*, 2002. 69(5), pp. 495-506.
- [11] A.J. Trippe, "Patinformatics: Tasks to tools," *World Patent Information*, 2003. 25(3), pp. 211-221.
- [12] S.Morris, et al., "DIVA: a visualization system for exploring document databases for technology forecasting," *Computers & Industrial Engineering*, 2002. 43(4), pp. 841-862.
- [13] R.S. Campbell, "Patent trends as a technological forecasting tool," *World Patent Information*, 1983. 5(3), pp. 137-143.
- [14] C. Lee, et al., "A stochastic patent citation analysis approach to assessing future technological impacts," *Technological Forecasting and Social Change*, 2012. 79(1), pp. 16-29.
- [15] H. Chen, G. Zhang, J. Lu, and D.Zhu, "A Two-Step Agglomerative Hierarchical Clustering Method for Patent Time-Dependent Data," *The 7th International Conference on Intelligent Systems and Knowledge Engineering (ISKE2012)*, Beijing, China, in press.
- [16] C. Christodoulos, C. Michalakelis, and D. Varoutas, "Forecasting with limited data: Combining ARIMA and diffusion models," *Technological Forecasting and Social Change*, 2010. 77(4), pp. 558-565.
- [17] E. Keogh, et al. "An online algorithm for segmenting time series," *2001 IEEE International Conference on Data Mining*, California, USA, 2001.
- [18] R.J. Watts, and A.L. Porter, "R&D cluster quality measures and technology maturity," *Technological Forecasting and Social Change*, 2003. 70(8), pp. 735-758.
- [19] P.L.Chang, C.C. Wu, and H.J. Leu, "Using patent analyses to monitor the technological trends in an emerging field of technology: a case of carbon nanotube field emission display," *Scientometrics*, 2010. 82(1), pp. 5-19.

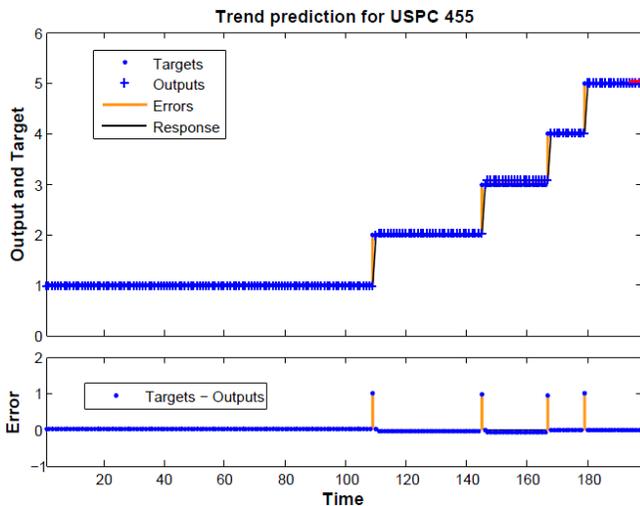


Figure 6. The prediction result of USPC 455 time series

V. CONCLUSION AND FUTURE WORK

Technical advances and accumulation of intellectual property of modern society will keep being emphasized in future. Automatic tools for assisting decision makers in a technology analysis domain will be in great demand. In existing theoretical research, the technology intelligence concept usually covers a very broad scope, although some data mining approaches have been used for patents analysis and technology forecasting, and key procedures like future trend forecasting still depend on expert-based qualitative methods that are normally expensive and time consuming. In addition, content-based technology intelligence that is built on text mining techniques usually fails to cover the time-related knowledge, which is also a part of decision makers' requirement. Under such circumstances, time-related property of patent data needs to be taken into consideration.

As a remedy, this paper presents an innovative framework of time-series-based TI by first applying time series analysis to the system architecture. The trend prediction component we proposed overcomes the problem that the prediction of values in future time intervals is insufficient to support TI construction. In addition, based on explicit patent properties and unknown patterns learned from the patent data, this research considers the two types of properties of patent data at the same time and combines the "trend" and "content" of patent knowledge for technological strategy support and future developmental planning. The framework proposed focuses on quantitative representation that comes from patents observation, which reduces expert participation to a great extent. Moreover, it improves and extends current technology intelligence concepts and frameworks architecture.

Although a time-series-based framework for technology intelligence is proposed in this paper that enables TI to be more effective when dealing with trend prediction requirement, more work is required. The learning and clustering modules of the framework need to be adjusted by series experiments. Hence, there will still be a lot of system implementation works addressed in future research.