Technological Forecasting & Social Change 119 (2017) 80–97

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Using the data mining method to assess the innovation gap: A case of industrial robotics in a catching-up country

Dejing Kong [a,b,c], Yuan Zhou [b,*], Yufei Liu [d], Lan Xue [b]

[a] School of Modern Post, Beijing University of Posts and Telecommunications, Beijing, China
[b] School of Public Policy and Management, Tsinghua University, Beijing, China
[c] The CAE Center for Strategic Studies, Chinese Academy of Engineering, Beijing, China
[d] College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, China

ARTICLE INFO

Article history:
Received 20 September 2016
Accepted 20 February 2017
Available online 18 March 2017

Keywords:
Innovation gap assessment
Support vector machines-based classifier
High-quality patents
Industry robot
China

ABSTRACT

It is critical for "catching-up" countries to narrow innovation gaps with developed countries by developing emerging industries. This research introduces a data-mining based method to systematically assess the national innovation gap that is specifically for emerging industries. The method examines the five key attributes of emerging industries, including the ownership of platform technologies, globalization intention, international knowledge position, university-industry linkage, and cross-disciplinary technology development. In particular, this method combines data-mining with experts' knowledge to build patent-training examples, and then uses a support vector machine-based classifier to single out all high-quality patents for each innovation attribute. Based on the selected high-quality patents, the authors utilize a factorial design analysis to systematically evaluate the innovation gap between countries. This method can significantly reduce measurement bias of traditional single patent indicators. In addition, it also can robustly adjust measuring weights in response to the specifics of each innovation attribute, while traditional multi-attribute evaluation methods cannot. As a result, this research empirically shows that China' industrial robot sector has apparent innovation gaps compared to developed economies, specifically in university-industry linkage, cross-disciplinary competence, and globalization intention, and this calls for the attention of policy makers and industrial experts.

1. Introduction

Innovation is necessary for "catching-up" countries (Fan, 2006). Backward countries – at different times – have managed to narrow the gap in innovation between themselves and the frontier countries, and we call it "catch up". Studies have noted that innovation is a major stimulus for national economic growth in industrial, newly industrialized, and developing economies (Archibugi et al., 1991; Ernst and Kim, 2002; Guan and Chen, 2012; Kim, 1980; Pavitt and Walker, 1976). Further, an effective diffusion of innovation is vital for the economic development of many countries' operating periods under different social and economic systems (Guan et al., 2005). Every country is a beginner in the newly emerging techno-economic paradigm, and innovation capability can serve as a cause for catching up (Schumpeter, 1942). Latecomers can catch up with more advanced countries by leap-frogging, or direct innovation at the technological frontier. Shortening the innovation gap with developed countries is meaningful, as well as achieving leaps in development, by developing emerging industries to facilitate this catch-up (Perez, 2010).

Multiple methods exist to assess innovation and innovation gaps across entities, using a variety of attributes as noted in Table 1. Innovation is a comprehensive result of multiple factors, and it is difficult to evaluate innovation based only on objective data. Thus, the case study is a popular method to illustrate innovation capability and the gap between latecomers and frontiers (Hobday, 1998; Fan, 2006; Fu et al., 2011; Choung et al., 2014; Rogo et al., 2014; Gao, 2015; Ernst, 2015). A survey-based quantitative analysis is another effective method to evaluate innovation capability and this gap (Anderson et al., 2013; Forés and Camisón, 2016; Guan and Yam, 2015; Vecchi and Brennan, 2009; Wu et al., 2016; Zehir et al., 2015). The results through case

Abbreviations: AHP, analytic hierarchy process; CN, China; CTD, cross-disciplinary technology development; DWPI, Derwent World Patents Index; EPO, European Patent Office; GE, Germany; GI, globalization intentions; HighValue, high-quality patent counts; IKP, international knowledge positions; INPADOC, International Patent Documentation Center; IPC, International Patent Classification; JP, Japan; JPO, Japan Patent Office; KR, South Korea; LDA, Latent Dirichlet Allocation; OPT, ownership of platform technologies; PCT, Patent Cooperation Treaty; Percentage, percentage of high-quality patents in all patents in a country; SVM, support vector machine; TI, Thomson Innovation; UIL, university-industry linkages; US, the United States; USPTO, United States Patent and Trademark Office.

* Corresponding author.
E-mail addresses: zhou_yuan@mail.tsinghua.edu.cn, zhouyuan.joseph@gmail.com (Y. Zhou).

http://dx.doi.org/10.1016/j.techfore.2017.02.035

Table 1
Attributes to assess innovation for the emerging industry.

Assessing attribute	Studies by case study on innovation			Studies by survey on innovation			Studies by econometrics on innovation		
	Fan (2006)	Rogo et al. (2014)	Gao (2015)	Guan and Yam (2015)	Forés and Camisón (2016)	Wu et al. (2016)	Corrocher et al. (2003)	Fu and Yang (2009)	Liu and Zhi (2010)
Ownership of platform technologies	*		*	*				*	*
Globalization intention						*	*		
international knowledge position		*	*		*				*
University-industry linkage	*	*	*		*		*	*	*
Cross-disciplinary technology development		*					*		

Assessing attribute	Studies by econometrics on innovation		Studies by bibliometrics on innovation						
	Li et al. (2016)	Castellacci and Natera (2016)	Porter and Detampel (1995)	Hung and Chu (2006)	Srinivasan (2008)	Bekkers and Martinelli (2012)	Wu and Mathews (2012)	Ávila-Robinson and Miyazaki (2013)	Li et al. (2016)
Ownership of platform technologies						*	*	*	*
Globalization intention	*	*							
International knowledge position		*	*			*	*	*	*
University-industry linkage			*	*	*		*	*	*
Cross-disciplinary technology development				*	*		*		

study or survey methods can be easily affected by the selection of cases and interviewees. Another type of quantitative method is based on patents, journal publications, news, and economic data that focus on innovation capability and diffusion efficiency (Ávila-Robinson and Miyazaki, 2013; Castellacci and Natera, 2016; Fu and Yang, 2009; Gu et al., 2016; Li et al., 2016; Liu and Zhi, 2010; Mellor and Hyland, 2005; Oura et al., 2016; Shao and Lin, 2016; Wu and Mathews, 2012).

Patent data can effectively indicate innovation performance, including product, process, and technology innovation, which is especially more accurate than such alternative measures as "new product" sales (Acs et al., 2002; Choi et al., 2011; Fu, 2008; Hong and Su, 2013; Jaffe et al., 1993; Usai, 2011; Wang and Lin, 2013). Previous research has always selected one single indicator, or a package of single indicators, to indicate various countries' innovation capabilities based on patent data, such as citations (Guan and Gao, 2009; Harhoff et al., 2003; Liu and Zhi, 2010), the number of publications (Fu and Yang, 2009), claims (OuYang and Weng, 2011; Tong and Frame, 1994), and the number of countries in which the patents are filed (Ernst and Omland, 2011; Harhoff and Hoisl, 2007; Meyer et al., 2011), among others. These indicators are always easily obtainable patent features, and they assume that the patent's quality or quantity can be presented by one indicator in one dimension. However, quality is a comprehensive effect achieved through different patent features. Multi-criteria methods also exist to indicate innovation, such as the analytic hierarchy process (AHP), and these can compare different countries' patent portfolios using different indicators' weights, measured by expert assessment. However, the real case is complicated, caused by multiple technology categories of owned patents, multiple countries that have prioritized patents, and multiple time periods in which the patents were published. The weights differ under various conditions, and the mass data characteristics cannot be adequately and comprehensively processed in batch mode.

Therefore, this research proposes a new method, support vector machines (SVMs), to identify high-value patents and assess innovation gaps between different countries based on high-quality patents. This is a popular and effective supervised-learning method, which asks a machine or algorithm to learn from the training sets for patent classification (Venugopalan and Rai, 2015). Experts can select a set of training examples in one classification (including two categories: one positive and one negative category) based on expert knowledge, which is similar to using more complex composited indicators. Not all high-quality patents require expert selection, while all of the patents in the positive category can be guaranteed as high quality. Similarly, all of the patents in the negative category can be guaranteed as low quality. When an expert identifies whether a patent is high quality, he considers multiple patent features, with different resulting contributions. However, no fixed weights exist for different features, such as traditional AHP methods. The contributions of features regarding high-quality patents' identification are more flexible to reflect experts' knowledge. Additionally, SVMs have the absolute advantage in handling massive amounts of data.

This research will use SVMs to assess innovation gaps between late-development and developed countries, and use a factorial design analysis to investigate the direction in which more investments are necessary (Beck-Broichsitter et al., 2012; Macdonald, 2011) in an empirical study of industrial robot innovation. Industrial robotics is a compelling, emerging, and important enabling technology, with radical novelty and relatively fast growth, coherence, prominent impacts, and uncertainty (Rotolo et al., 2015). This industry receives increasing attention with manufacturing developments, especially when it proposed that integrating artificial intelligence, robotics, and digital manufacturing technology was revolutionizing manufacturing. China is a late-developing country in industrial robotics compared with the United States, Japan, and Germany. This study's goal is to evaluate the innovation gap between leading industrial robotics countries and late-development countries by integrating SVMs based on patent data to address the following questions:

Q1: Does the SVM-based method provide reliable and valid innovation assessment results?
Q2: How large are the industrial robotics innovation gaps among the United States, Japan, Germany, and China?
Q3: How can the industrial robotics innovation gap change for China?

National innovation is a comprehensive performance related to multiple attributes. The authors measure an emerging industry's innovation gap between late-developing and developed countries through an assessment from five perspectives: the ownership of platform technologies, globalization intentions, international knowledge positions, university-industry linkages, and cross-disciplinary