Routledge
Taylor & Francis Group

Technology Analysis & Strategic Management

ISSN: 0953-7325 (Print) 1465-3990 (Online) Journal homepage: https://www.tandfonline.com/loi/ctas20

Evaluating the competitiveness of enterprise's technology based on LDA topic model

Xuefeng Wang, Xuemei Yang, Xinglu Wang, Mengge Xia & Jiayun Wang

TECHNOLOGY ANALYSIS & STRATEGIC MANAGEMENT
https://doi.org/10.1080/09537325.2019.1648789

Routledge
Taylor & Francis Group

Check for updates

Evaluating the competitiveness of enterprise's technology based on LDA topic model

Xuefeng Wang, Xuemei Yang, Xinglu Wang, Mengge Xia and Jiayun Wang

School of Management & Economics, Beijing Institute of Technology, Beijing, People's Republic of China

ABSTRACT

With the advent of the knowledge economy, technology has become the foundation of advancement for many enterprises. To maintain a lasting competitive edge, enterprises must accurately determine the competitiveness of particular technologies. However, emerging innovations are becoming more and more complex, and interdisciplinary trends complicate matters even further. In the competitiveness evaluation based on patent, traditional patent classifications are both loose and time-sensitive. In this paper, we constructed the evaluation model of enterprise's technology competitiveness based on the technological topics generated by a latent Dirichlet allocation (LDA) topic model. LDA topic model is able to classify technologies into narrower categories and can, therefore, provide rich information on the competitive landscape of a field. Two indexes are used to determine the technological competitiveness of an enterprise in the model – a specialisation index and a diversification index. At the same time, we explore the distribution of enterprises with different technological topics through the relative share, the technology's appeal, and the competitive advantage that technology might give an enterprise. The empirical study on intelligent connected vehicles validates the model, and the results provide theoretical support for developing R&D strategies and/or making investment decisions.

ARTICLE HISTORY
Received 22 November 2018
Revised 23 May 2019
Accepted 16 July 2019

KEYWORDS
Technology competitiveness; patent mining; patent portfolio; topic model

1. Introduction

In today's fiercely competitive world, enterprises must establish competitive advantages to conduct business with a healthy growth path, especially advantages based on technology. For most technology-based enterprises, innovations with strong market appeal run through the entire value chain. They rely on the uniqueness and competitiveness of their technology to maintain product novelty and efficiency, promote organisational change, expand market share, and thus remain invincible in the face of fierce domestic and international competition (Koellinger 2008). And small differences in these technologies can make a big difference to the long-term performance and competitive position of the enterprise (Barney 1991).

However, since the advent of big data and the knowledge economy, the complexity of emerging technologies is increasing (Breschi, Lissoni, and Malerba 2003; Garcia-Vega 2006). Thus, for many firms, the current focus is on controlling intellectual property rights. To this end, Brockhoff (1992) used patent portfolio data, i.e. the suite of patents held by a firm, to analyse the technology strategies of various companies and, in particular, the competitiveness of their technologies. Although patent data has certain limitations for evaluating technological competitiveness, patents still have the advantage of homogeneity, detail, accuracy, availability, and lower costs. Scholars generally believe that patent data is an important and rich source of empirical data for examining

CONTACT Xuefeng Wang wxf5122@bit.edu.cn

technology-related issues, such as the distribution of technology or competition within a technology field (Cho and Park 2015; Cui et al. 2018).

Yet, when comparing and contrasting a range of technologies, it is necessary to classify the technology according to some criteria. Currently, most researchers use the International Patent Classification (IPC) system or similar. These classification systems are relatively subjective – a feature that underpins their main limitations (Venugopalan and Rai 2015) and means that patent data are often unsuitable to fully meet the needs of enterprise planning (Lai and Wu 2005). Herrero et al. (2010) point to the need for frequent updates. Further, in a review of the managerial challenges with analysing patent data, Nakamura et al. (2015) highlight the high cost of data collection and the limited success in producing practical results (see also Kajikawa, Abe, and Noda 2006). Through expert interviews, they conclude that even though patent data is a relevant decision-making tool for practitioners, its usefulness is hindered by the inherent limitations of classification-based metrics.

More recently, emerging technologies, such as machine learning, have led to revived interest in developing new and different methods for mining patent data, such as semantic topic mining and topic clustering. For example, Yoon and Kim (2012) developed a content analysis system for patent data based on subject-action-object phrase structures and topic clustering techniques to construct a dynamic map of technological competition trends in R&D activity. Suominen, Toivanen, and Seppanen (2016) used an unsupervised latent Dirichlet allocation (LDA) model to classify technology topics based on the full text of patents for a number of enterprises over time to create an overall view of the patents in an entire industry. The results were used to predict future trends. Based on the topic recognition, Ma et al. (2017) analyse the competitiveness of the organisation based on the three-layer probability distribution generated by the topic model. While these techniques represent progress, more effective text mining methods are needed to further evaluate the competitiveness of an enterprise's technology using patent data.

Coupled with unsupervised machine learning, LDA topic models fully automate the process of producing a probability distribution of the topics in a corpus (Patel and Pavitt 1997). Hence, in this paper, we use an LDA topic model to identify technologies as topics and then build a three-layer enterprise-patent-topic probability distribution. The model is constructed based on two indexes of a firm's patent portfolio – one that measures specialisation and another that measures diversification. Once built, the model can be used to evaluate the competitiveness of each company and each technology. Further, the results can be used to identify the existing development strategy of each company and provide a theoretical basis for implementing a new or modified R&D strategy appropriate for meeting the firm's future business goals.

The rest of this paper is organised as follows. Section 2 provides a literature review of the definitions and methods associated with evaluating technological competitiveness. Section 3 presents the research framework and construction of the evaluation models. Section 4 provides a case example to demonstrate the evaluation and analysis process and illustrate how the model can be used for practical applications. Section 5 concludes the paper with the implications and limitations of this research and our future research directions.

2. Literature review

2.1. Definition of enterprise technology competitiveness

A unified definition of an enterprise's technological competitiveness has not yet been formed. In keeping with the competency school of thought, Prahalad and Hamel (1990) believe that a firm's sustainable competitive advantage and business performance are derived from the core competitiveness of the company and that technological ability is the key contributing factor to core competitiveness. Technological competitiveness is reflected in a firm's ability to innovate, coordinate, and integrate multiple production skills and technology flows. In the same school, Mayindi and

Kachienga (2008) believes that improving capability is the key to improving technological competitiveness and, in turn, gaining a market advantage. Jian (2006) believes that technological competitiveness comes from relevant capabilities owned by enterprises, such as technological capabilities, technological innovation capability. Leonard-Barton (2010) takes a broader view, proposing that the skills and production capabilities of a firm's employees *are* the enterprise's technology. Grindley (2010) argues that technological advantages usually depend on an organisation's structure and good organisational skills help companies turn technological advantages into competitive advantages.

Wernerfelt (1984), a representative scholar of the resource school, believes that technology competitiveness of enterprises is a unique collection of technological resources for enterprises. When studying the technology competitiveness of enterprises, these scholars generally believe that the quantity and quality of the technological resources within an enterprise are its main source of technological competitiveness. These are also key factors for gaining a dominant competitive position.

In the evaluation model, we define enterprise's technological competitiveness according to the combined viewpoints of the resource and competency schools. Competitiveness stems from the collection of technological resources within a company that cannot be easily imitated by others, and the ability of the company to leverage a competitive advantage with that technology.

2.2. Evaluating technological competitiveness

In general, technological resources are highly specialised and, therefore, difficult to categorise. As such, they are highly knowledge-intensive intangible assets. Patents are often the commercial realisation of concentrated technical knowledge and an important resource for success in long-term development strategies. Beyond being a rich source of information and the output of an enterprise's innovation activities, they also often form the inputs for the next stage of a technology's development. Therefore, many scholars have evaluated technological competitiveness across enterprises using patent analysis. Many scholars have since used data mining technology to conduct extensive patent analyses to explore the technological activities and capabilities of enterprises. Pavitt (1985) studied the relationship between patent data and the innovation activity of particular companies based on theories of economics and bibliometrics. Narin, Elliot, and Ross (1993), the father of patent measurement, argues that patent analysis is an effective means of evaluating the technology competitiveness of an enterprise, asserting that patent data is the most effective indicator for describing and measuring the technological competitiveness of an enterprise. Topic model has been recently applied to patent data in a number of studies. Venugopalan and Rai (2015) used a topic-based approach to analysing the structure of patent data. Their analysis used the contextual frame of knowledge spillovers, resulting in the use of patent abstracts and claims as the basis of analysis.

At present, there are several main methods for evaluating the technological competitiveness of an enterprise using patent data: index systems, patent portfolio analysis, and patent network and matrix evaluations. The index system is the earliest method. Cantwell and Fai (1999) used relative technological advantage to characterise competitiveness, while Fai (2005) used granted patents. Yu and Lo (2009) measured competitiveness with a combined set of indicators including relative advantage, patent activities, citation rates, relative citations, etc. Eck and Waltman (2010) used the patent application category, authorisation class, patent maintenance, patent scope, and patent citations to construct an evaluation system for enterprise competitiveness. Principal component analysis was used to weight each index.

Brockhoff (1992) was the first to propose patent analysis on a company's full portfolio in 1991. Ernst (1998) subsequently developed four patent portfolio analysis models at different levels – most commonly at the technology and enterprise levels. Wang, Lo, and Liao (2015) focused on improving patent indicators to better evaluate and analyse technological competitiveness in various fields.

Patent network and matrix evaluations rely on citation relationships, co-occurrence relationships, or cooperation relationships between patents to construct a corresponding technological network to evaluate the technological competitiveness of enterprises. For example, Ramani and Looze (2002) used various patent features, such as the patent holder, patent applications, and IPC classifications to measure competitiveness from the density and centrality of patents.

However, although using indicators as evaluation metrics provides a wide scope for assessing specific aspects of technological competitiveness, most demand some kind weighting and these weightings can be subjective or based on inaccurate information. For instance, the majority of citations in patents are included by the examiner and often reflect technological priorities rather than a close relationship to patent's innovation. Therefore evaluation methods based on citations, such as the network and matrix method, may be drawing on fundamentally flawed information. Whereas, evaluation methods based on an enterprise's patent portfolio can use different indexes at different evaluation levels to conduct multi-dimensional comparisons. As such, it is easier to identify competitors and analyse competition with these methods. Given the importance of a competitive advantage to most technology companies, the results of this type of analysis are a more rational basis for allocating resources and formulating innovation strategies. Further, the evaluation results from a patent portfolio analysis are easier to illustrate in a visible and understandable way.

3. Methodology

The LDA topic model for evaluating the competitiveness of enterprises based on the technologies they hold in this paper is divided into four main steps: data acquisition and pre-processing, identifying technology topics, constructing the competitiveness model, and evaluating competitiveness. The model's framework is shown in Figure 1.

3.1. Data acquisition and pre-processing

The framework is based on patent information as its data source. While any patent database could be used, we retrieved our corpus from Derwent Innovation (DWPI). Given that patent documents inherently contain unstructured textual data, investigating all possible semantic phrases to define a technology topic would be very challenging (Marikkannan, Marikkannan, and Kannan 2008). Abstract-DWPI contains a wealth of patent information, including a summary of the patent's contents prepared by field experts at Derwent. Hence, using the patent's summarised claims combined with Abstract-DWPI as the source data better captures the complexity of the innovation and the enterprise's knowledge.

Once the corpus is assembled, each patent document is converted into structured data using keyword vectors. Words that are closely associated with a specific technology and appear frequently in patent documents could also be used as keywords. A keyword vector comprises the keyword and the frequency with which it appears in the patent. We used ITGInsight software to clean the data, extract the keywords, and create the keyword vectors (Zhang et al. 2013). The topic vocabulary was constructed manually.

3.2. Identifying the topics

LDA topic models were first proposed by Blei, Ng, and Jordan (2003). These models produce a three-layer Bayesian probability distribution based on probabilistic latent semantic indexing (pLSI). Essentially, this approach is a document generation probability model. Our LDA model consists of a three-tier structure of words, topics, and documents. The basic idea is to treat documents as a mixture of their implicit topics with each topic appearing as a probability distribution of the words related to the topic. LDA can be used to identify potential topic information in large document sets or corpora.

The LDA topic model is widely used in subject recognition (Hu, Shu, and Tian 2014), subject evolution analysis (Guan, Wang, and Fu 2016), patent content analysis (Wang et al. 2015), technology topics evaluation (Ma et al. 2017), and scientific text classification (Lee, Han, and Sohn 2015). The

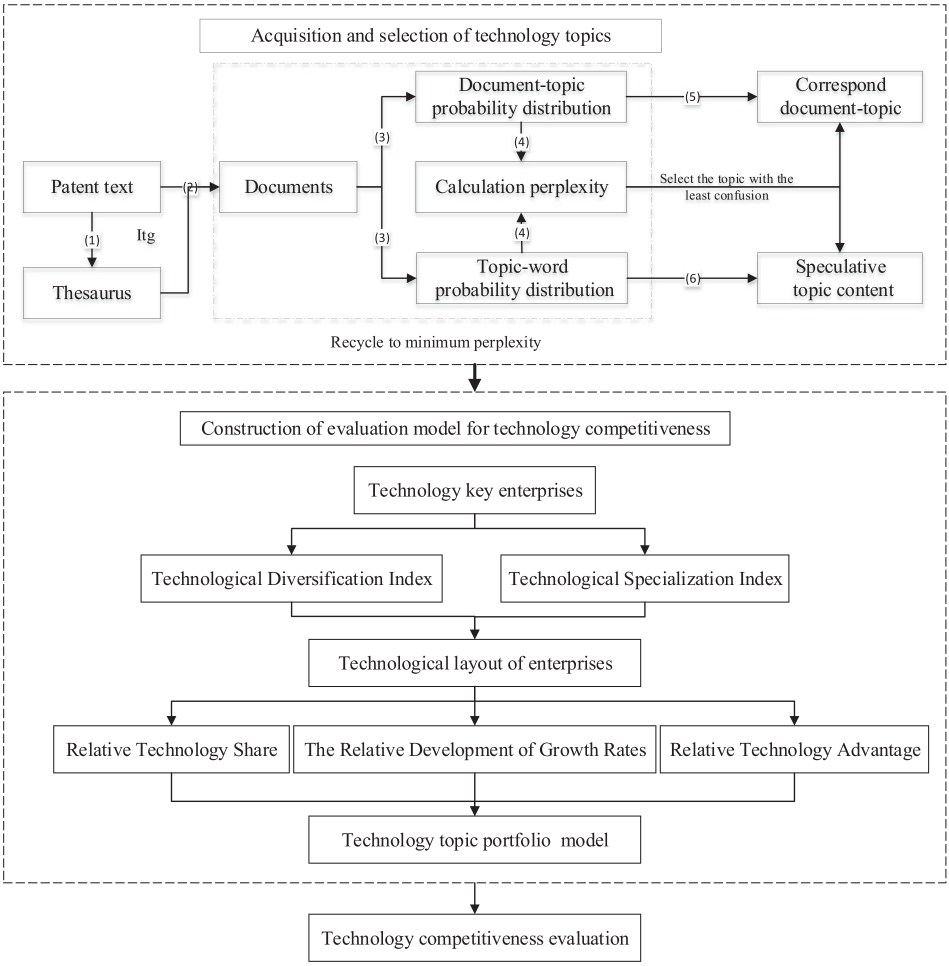

Figure 1. Framework of the enterprise's technological competitiveness model.

above applications need to determine the number of technological topics. Following Blei and Lafferty (2006), we use perplexity as the criterion for evaluating the quality of the model. The perplexity is calculated as follows:

$$\text{perplerity} = e^{-\sum \log(p(w))/N}, \tag{1}$$

where $p(w)$ represents the probability of occurrence of the characteristic word w in the patent, the formula for $p(w)$ is as follows:

$$p(w) = \sum zp(z|d)*p(w|z). \tag{2}$$

The optimal number of topics depends on the model with the least perplexity. The smaller the perplexity, the better the predictions generated by the model.

LDA topic models also contain hyperparameters, and these need to be defined. In a previous study, Griffiths (2004) obtained good model quality and topic clustering results with the hyperparameter settings $\alpha = 50/K$ and $\beta = 0.01$. Therefore, we used those settings for our analysis. For a detailed explanation of the calculation process and algorithms, we refer readers to Blei and Lafferty (2006) and to Yau et al. (2014).

As mentioned above, the appropriate number of topics depends on the perplexity, which is determined by defining a threshold. However, it is important to standardise this threshold based on the size of the enterprise. The final set of topics is inferred from the keyword vectors that represent the patents' contents.

3.3. Constructing the model

Model construction is based on both a diversification index and a specialisation index. The Herfindahl index is a commonly used measure of diversification. A transformation of the Herfindahl index has previously been applied to patent portfolios as a reflection of the technological diversification of a company. Technological diversification represents the breadth of the technology topics covered by the company's patents. This technological diversification (TD) index is defined as

$$\mathrm{TD} = 1 - \sum_{i=1}^{n} \left(\frac{f_i}{f}\right)^2, \tag{3}$$

where n represents the total number of technology topics, f_i represents the weighted amount of patents containing the subject category i in the company, and f represents the total number of patents in the company. The greater the diversity value TD, the more diversified the company's technology fields and the wider the coverage of technology topics.

Technological specialisation (TS) refers to the degree of centralisation of technology topics in the patents the enterprise owns. The specialisation index used in this model was proposed by (Porter et al. 2007) as follows:

$$\mathrm{TS} = \frac{\sum (f_i \times f_j \times \cos(t_i \times t_j))}{\sum (f_i \times f_j)}, \tag{4}$$

where t_i and t_j indicates that the patents relate to the category of the topic, $\cos(t_i \times t_j)$ indicates the degree of association between category t_i and category t_j. The larger the specialisation value TS, the stronger the professionalism and the higher the competitive advantage the company holds.

In order to further explore the distribution of enterprises with different technological topics we used the patent portfolio method (Ernst 1998) with the extensions developed by Brockhoff (2002) and Wang, Lo, and Liao (2015). However, we made some improvements to capture and assess relative share, technological appeal, and relative advantage as part of the evaluation. To normalise the quantitative differences between the number of topics in each patent, the LDA model applies a weighting and then generates the probability distribution. The specific calculation formulas for each index follow.

(1) *Relative technology share* (*RTS*)

This is a measure of the gap between a firm's patent rankings in a given technology field and that of the benchmark company in that field. The formula for calculating the relative technical position is

$$\mathrm{RTP} = \frac{P_{ij}}{p_{iB}}, \tag{5}$$

where p_{ij} denotes the number of patents for company j in technology field i, p_{iB} as the weighted number of patents in the i technological field of the benchmark company. RTP values fall in the range [0, 1]. Higher values indicate a higher technological position in the field. The maximum value of 1 represents the benchmark company.

(2) *The relative development of growth rate* (*RDGR*)

This metric represents the growth rate of patents granted within a technology topic which also called technological attractiveness. Studies have shown that technology topics with high patent growth rates will be more attractive in the future than technology topics with lower patent growth rates. The formula for technical attraction is

$$\mathrm{RDGR} = \frac{\text{Growth rate of patent grants for a single technology}}{\text{Growth rate of patent grants for all technologies}} \tag{6}$$

(3) *Relative technological advantage* (*RTA*)

This is a standard measure of the strength of a technology. An RTA > 1 represents a relatively high technology ability. Vice versa, it represents a relatively low technology ability. The technological advantage value can be used to judge the enterprise's technological ability in a topic. The formula is

$$\mathrm{RTA}_{ij} = \frac{p_{ij}/\sum_j p_{ij}}{\sum_j p_{ij}/\sum_{ij} p_{ij}}. \tag{7}$$

3.4. Evaluating competitiveness

The first step in evaluating the competitiveness of a technology is to determine the distribution of a firm based using the specialisation and diversification indexes. Then, competitiveness is analysed at the topic level.

A two-dimensional matrix is constructed using these two indexes, as shown in Figure 2. To avoid the influence of extreme values, the four quadrants are divided by the median of the two indicators. Quadrant 1 is low specialisation-low diversity. Here, enterprises tend to be start-ups. Companies in Quadrant 2 own multiple technologies, but not enough to form technological barriers. Quadrant 2 is high specialisation-high diversity, where enterprises are generally technology leaders. These enterprises have a technological advantage in a number of technologies. Quadrant 4 contains companies with a single or very limited number of technologies. Enterprise advancement is limited but the technologies form high technological barriers.

The bubble chart shown in Figure 3 takes the relative technology share as the abscissa and technology attractiveness as the ordinate. The four quadrants are divided according to the relative development of growth rates of the benchmark, i.e. 1, and the average of the relevant technology shares. The size of the bubble size represents the relative technical advantage. Within a technology topic, it more valuable to have a higher technology share and a higher attractiveness.

4. Case study

4.1. Research subject selection

As the automobile industry becomes more internationalised, industrial boundaries are becoming increasingly blurred and emerging technology companies are entering the supply and value

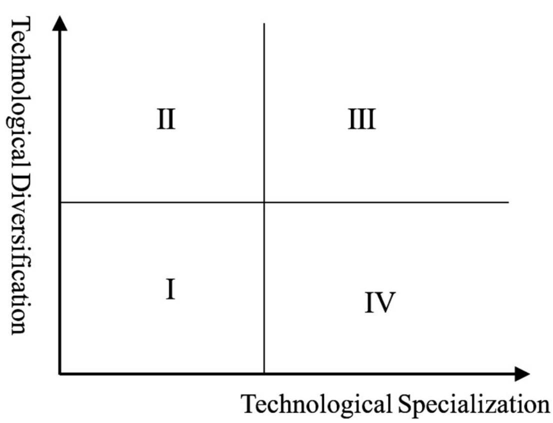

Figure 2. The bubble chart of evaluation.

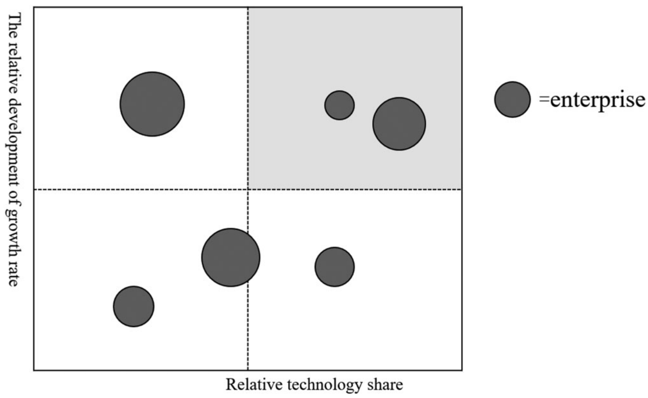

Figure 3. The evaluation bubble chart.

chains. Traditional enterprises now compete with emerging enterprises and the global automotive industry ecology is being reshaped. We chose the field of intelligent, connected vehicle technology for this empirical study because this field is emerging, novel, and complex. Intelligent connected vehicles (ICVs) are equipped with advanced in-vehicle sensors, controllers, actuators, and other devices to form a complex system of car networking and smart cars. Modern communication and network technologies integrate to exchange and information between vehicles, people, roads, and situations. In future, this new generation of cars will be used for autonomous driving.

A search of DWIP retrieved 11195 patent records; 9849 were relevant. The top 20 enterprises according to number of patents were selected as subjects for evaluation: 12 were traditional vehicle manufacturers and 8 were emerging enterprises (marked as emg). The research object and the number of patents it holds are shown in Table 1.

4.2. Identifying the technology topics

Following the method outlined in Section 3, we calculated the topic perplexity (Equations (1) and (2)), shown as a curve in Figure 4. Despite the clustering accuracy that comes with a smaller perplexity value, the number of topics was too large to reduce dimensionality. Therefore, 300 topics were chosen to smoothen the curve.

To test the reliability of the model, we calculated the similarity between topics according to the word/topic probability distribution. The result of 0.0014 indicates a low level of redundancy. Through a series of tests, we set the topic threshold to 0.02. Each patent contained a number of topics, which suggests that the 300 technology topics had high coverage across all patents. The topic-patent correspondence probability for the 20 enterprises was standardised on this basis.

Table 1. Intelligent connected vehicle enterprises and patents.

Company	Abbrev	No. patents	Company	Abbrev	No. patents
GM Global Technology Operations Inc.	GENK	655	Nissan North America, Inc.	NSMO	90
Hyundai Motor Co., Ltd.	HYMR	328	Honda Motor Co., Ltd.	HOND	77
LG Electronics	GLDS(emg)	269	DaimlerChrysler AG	DAIM	69
Toyota Motor Eng & Mfg	TOYT	208	Robert Bosch GmbH	BOSC	67
Google	GOOG(emg)	183	Volkswagen AG	VOLS	67
Hyundai Mobis Co., Ltd.	HYMO	177	Fuji Heavy Industries Ltd.	FUJH	48
Ford Global Technologies LLC	FORD	154	Intel Co., Ltd.	ITLC(emg)	47
Marvel International Ltd.	MVLL(emg)	137	Chery Automobile Co., Ltd.	CHRA	47
Nippondenso Co., Ltd.	NPDE(emg)	109	Aisin AW Co., Ltd.	AISW(emg)	46
Samsung Electronics Co., Ltd.	SMSU(emg)	95	Mitsubishi Electric Co., Ltd.	MITQ(emg)	45

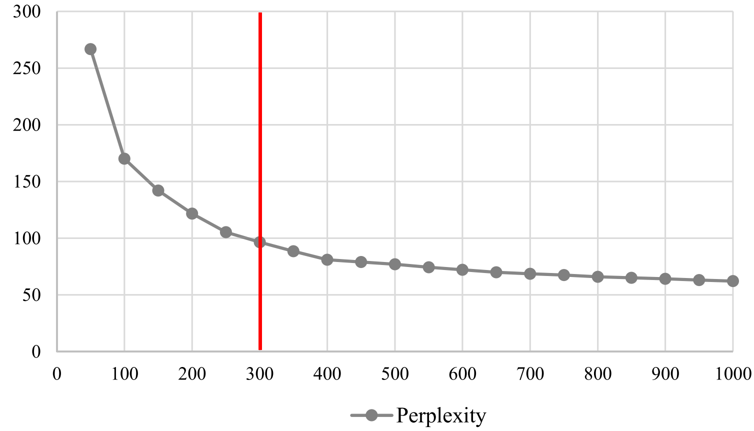

Figure 4. Topic perplexity curve.

4.3. Evaluating the technological competitiveness of ICV enterprises

4.3.1. Distribution of the enterprises

Figure 5 shows the number of topics and coverage for each enterprise. Due to the high number of topics, we have not listed them here.

The patent portfolios for each enterprise spanned at least 100 topics. GM Global Technology Operations Inc. (GENK) had the highest number of patents, but Hyundai Motor Co., Ltd. (HYMR) had a greater topic coverage. Companies with lesser coverage included the Intel Co., Ltd. (ITLC) and Aisin AW Co., Ltd. (AISW).

We then constructed the specialisation and diversification matrix based on the three-layer probability distributions produced by the model (Equations (3) and (4)). The competitiveness scatter plot is shown in Figure 6. From a broad perspective, all 20 enterprises ranked relatively high on the diversification index with 100 or more topics represented in each patent portfolio. This result is somewhat unsurprising since the intelligent connected vehicle industry is currently in its exploration stage. There are no real technological barriers yet, and the manufacturing chain still needs many technological advancements before it can fully develop. However, there were some significant differences between each enterprise. DAIM, FORD, GOOG and VOLS fell into Quadrant 1, i.e. start-ups with

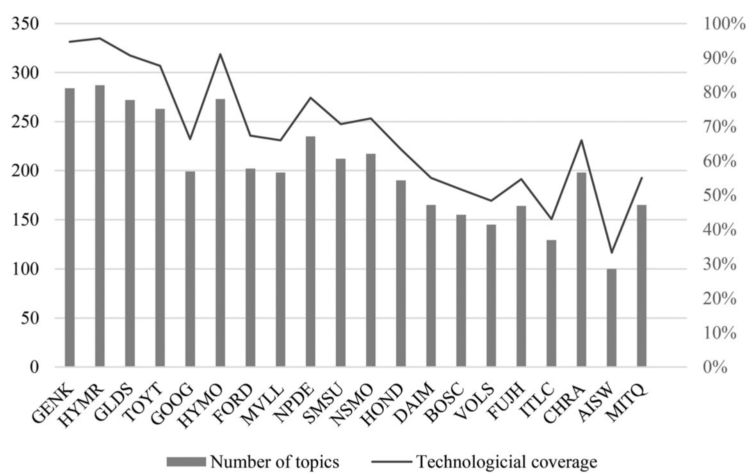

Figure 5. Number of topics and topic coverage for the top 20 enterprises.

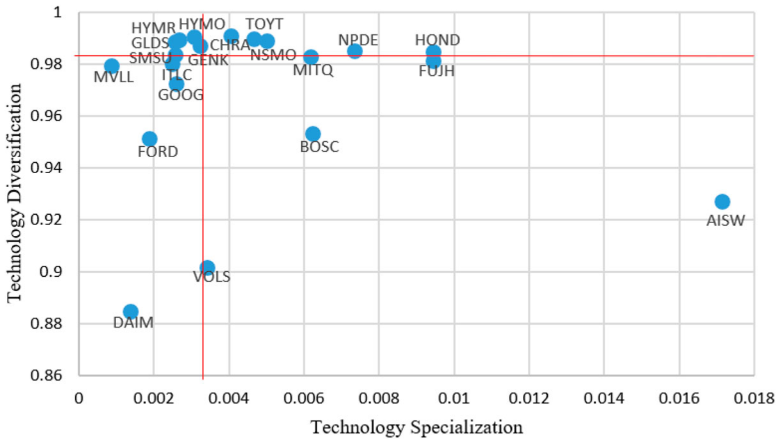

Figure 6. Competitiveness scatter plot of the top 20 enterprises.

both low specialisation and diversification. For example, DAIM entered the field in 2015, only two years before the intelligent connected vehicle industry began to take off. Most traditional vehicle manufacturers fell into Quadrant 2, particularly those with a high number of patents, signalling comprehensive development activity in this sector. Quadrant 3 contains a mixture of both traditional vehicle manufacturers and emerging enterprises, such as CHRA, TOYT, NSMO, NPDE, and HOND. It can be seen that both traditional vehicle manufacturing enterprises and emerging enterprises have a place in this field. Only two enterprises are squarely placed in Quadrant 4 – AISW and BOSC. Both are emerging enterprises and both specialise in early-stage artificial intelligence research and machinery manufacturing. Therefore, each holds a certain technological advantage in this field.

4.3.2. Competitiveness at the topic level

To further understand each enterprises' technological competitiveness, we analysed each enterprise at the topic level. While most topics were represented in the sample, we selected the 11 topics addressed by at least 19 or more of the companies.

We made an educated guess at the content of each of the 11 topics from the keyword information, as shown in Table 2.

Using Equation (4), we calculated the relative share, appeal, and technological advantage for each topic and generated a bubble chart, as shown in Figure 7. Recall that the abscissa represents the relative technology share, and the ordinate indicates attraction (the relative growth rate), the same type of technology has the same technology attraction. The size of the bubble represents the enterprise's technological advantage in a topic – the larger the bubble, the greater the technological advantage.

Table 2. The 11 technology topics covering a wide range of companies.

Topic	Technology area	Specific category
36	Intelligent decisions	V2X information fusion
51	Vehicle control	Intelligent induction brake protection system
109	Intelligent decisions	Traffic management decision system
123	Intelligent decisions	Vehicle diagnostic signal processing terminal
143	Intelligent decisions	Electronic chip, comparator
190	Wireless communication	Vehicle monitoring communication equipment
200	Wireless communication	Telematics systems
205	Vehicle controls	Vehicle distance measurement system
271	Wireless communication	Analog to digital conversion sensor
275	Data platforms	Road offset signal processing
287	Intelligent decisions	Long-distance image information processing system

Figure 7. The relative share for 11 technologies.

As the figure shows, there are obvious differences in technological appeal across the topics. There are also significant differences between the same technology held by different enterprises and different technologies of the same enterprise. The five topics below the baseline (the red line) signal that the intelligent connected vehicle industry is still in its preliminary stage. Innovation is gradually increasing, but there has not yet been a surge. Topic 190 (Driving assistance system) is the most attractive and, therefore, the most prominent.

Given that GM Global Technology Operations Inc. (GENK) and Hyundai Motor Co., Ltd. (HYMR) both hold a large number of patents, both companies hold a relatively high share in most topics. Marvel International Ltd. (MVLL) and Google (GOOG) hold absolute advantage in Topics 200 (Telematics systems) and 287(Long term evolution-vehicle). As a whole, the large traditional vehicle manufacturers hold the strongest technological advantages, but the power to innovate in emerging enterprises cannot be underestimated.

All companies in the sample hold at least one patent in Topic 51, which makes it a good representative example to evaluate the technological competitiveness of each enterprise. The keywords for this topic suggest it relates to vehicle path tracking controllers. The topic's appeal is less than 1, which indicates this technology is in a period of gradual development. The bubble chart for Topic 51 only appears in Figure 8.

HYMR has the largest technology share, while GENK, GLDS, TOYT, HYMO, NPDE, and HOND hold the strongest advantage. For example, NPDE only holds a small number of patents, but all are in key technologies. Hence, NPDE may be able to form technological barriers in a small scope. The results of this analysis also highlight that emerging enterprises are more focused on a single technology and

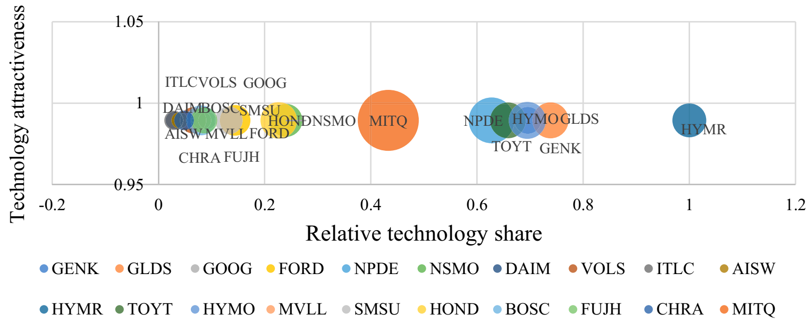

Figure 8 Enterprise competitiveness in Topic 51(vehicle path tracking controllers).

Table 3. The 12 technology topics of most concern to each company.

Topic	Enterprises	Technology area	Specific category
21	ITLC	Intelligent decisions	V2X information fusion
55	NSMO	Environmental awareness	Car image sensor chips
99	MITQ	Environmental awareness	Millimetre wave radar chips
130	CHRA	Intelligent decisions	Lane departure warning systems
138	SMSU	Vehicle control	Collaborative control systems
176	NPDE, HOND, AISW	Environmental awareness	Path planning systems
200	MVLL	Wireless communication	Telematics systems
201	TOYT, GOOG	Vehicle controls	Topological stream interaction
204	FUJH	Wireless communication	Obstacle detection, environmental surroundings warning systems
251	FORD, DAIM, BOSC, VOLS	Intelligent decisions	Adaptive cruise systems
257	GENK	Data platforms	Vehicle interconnection terminals
286	HYMR, GLDS, HYMO	Intelligent decisions	Night vision systems

have formed a strong competitive position through their potential to create technical barriers in key areas.

To further analyse competitiveness among the key technologies held by each enterprise, we selected the 12 technology topics of most concern to each company. We made an educated guess at the content of each of the 12 topics from the keyword information, as shown in Table 3.

Here, we analysed and compared competitiveness for each enterprise according to their core technologies. Although there is no comparability between different technological topics, this type of analysis can provide deeper insights into the broader competitiveness of each company.

Figure 9 shows the relative appeal versus share of each enterprise's core technology, i.e. the relative growth rate of the technology. A value of 1 is the reference line; therefore, a topic below that line has fading appeal and topics above the line have greater appeal. Topic 21 – V2X information fusion – has developed the fastest in recent years (RDGR = 5.67) and, given that this is a core technology for ITLC, it has a strong development advantage. However, it does not hold a large relative share. AISW, GOOG, CHRA, NSMO, and MVLL all play a leading role in their respective core technologies. Not only do these topics have high appeal, but these companies also hold an absolute technological advantage. In particular, CHRA holds the greatest technological advantage in lane departure warning systems – a technology that has the potential to create a technological barrier, making it more difficult for other companies to enter the field. The technologies with an RDGR of less than 1 include Topics 251 (Adaptive cruise systems), 257(Vehicle interconnection terminals), and 286 (Night vision systems). Although GENK has the largest number of patents and the largest technology share, vehicle interconnection terminals have been lagging behind in a state of stagnation. In a wave of rapid technological development, GENK failed to change its development strategy in time and has missed opportunities to develop new technologies. Conversely, companies with a small number of patents, such as AISW, ITLC, and MITQ, have greater technological advantages in topics with

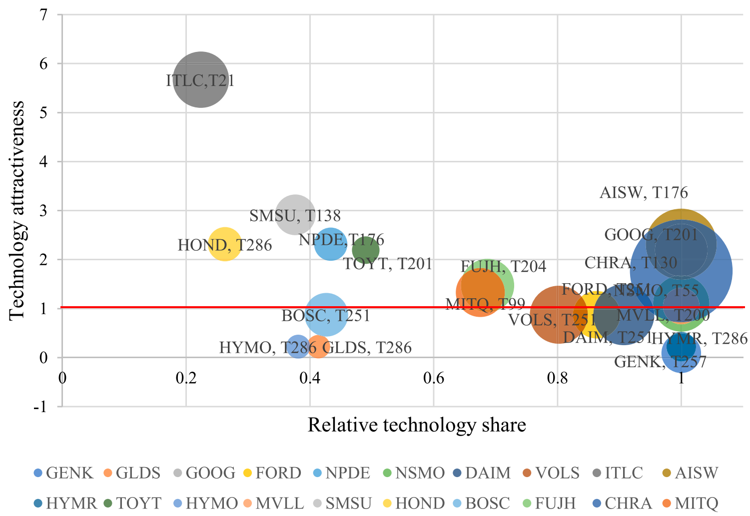

Figure 9. Enterprise competitiveness for key technologies.

higher appeal. Competition in Topics 251 and 286 is fierce, with HYMR leading the pack in adaptive cruise systems (Topic 251) and DAIM leading in night vision systems (Topic 286). Competitors in these fields will need to adjust their innovation strategies if they are to gain an advantage in future.

There are also interesting findings surrounding some of the emerging companies, such as AISW, GOOG, ITLC, and MITQ. Each has high levels of advantage in their core technologies and therefore have a high level of competitiveness.

The above analysis shows that: In general, the technology of traditional vehicle manufacturers is diversified, and the technology of emerging enterprises presents a trend of specialisation. The more patents a company has, the greater the share of technology in related topics, but it does not necessarily occupy an absolute technological advantage. Emerging enterprises are more focused on a single technology and occupy technological advantages in individual technology topics which have more technical attractiveness.

5. Conclusion

The IPC and other patent classification systems commonly used to analyse technological landscapes have some shortcomings. The classifications are coarse, cumbersome to break down manually, and often do not represent the latest technological innovations. Hence, we used an LDA topic model to the cluster topics found in patent documents and produce a three-layer probability distribution of enterprises, patents, and topics to assess the technological competitiveness of 20 enterprises in the field of intelligent connected vehicles. Each enterprise was ranked against two indexes – technological specialisation and technological diversity. While the concentration and dispersion of technologies is not an objective representation of an enterprise's competitiveness, it can show the differences and relative positions of each company. In this analysis, we identified the relative share, advantages, and appeal of the core technologies each enterprise in different specialist areas of the field is pursuing to provide insights into the competitive landscape between them. The results should be useful for informing future R&D strategies and investment decisions in intelligent connected vehicle technology.

This research has some limitations, which present opportunities for follow-up research. First, the specialisation and diversification scores for each enterprise were relatively concentrated, not scattered as expected. This is because the topics generated by the LDA model tended to be evenly distributed resulting in a more detailed clustering effect. Second, the large number of topics meant we could not analyse the full technological landscape of each enterprise. Hence, competitiveness was only assessed on a select number of commonly-shared topics. In subsequent research, we will further evaluate competitiveness across all topics. Lastly, the three-layer probability distribution was indirectly constructed by matching an enterprise with its patents and topics. However, an author-topic model would provide a more direct construction to yield different insights.

Acknowledgements

The findings and observations in this paper are those of the authors and do not necessarily reflect the views of the supporters. The authors would like to thank colleagues from the Beijing Institute of Technology.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work is supported by the General Program of National Natural Science Foundation of China [grant numbers 71774012, 71673024, 71373019] and the strategic research project of the Development Planning Bureau of the Chinese Academy of Sciences [grant number GHJ-ZLZX-2019-42].

Notes on contributors

Xuefeng Wang is professor at the School of Management and Economics, Beijing Institute of Technology, China. His specialty is technology innovation management, data mining and science and technology evaluation.

Xuemei Yang is a master student from the School of Management and Economics Beijing Institute of Technology. Her specialty is technology assessment and data mining.

Xinglu Wang is a master student from the School of Management and Economics Beijing Institute of Technology. Her specialty is data mining and emerging technology forecast.

Mengge Xia is a master student from the School of Management and Economics Beijing Institute of Technology. Her specialty is technology assessment.

Jiayun Wang is a master student from the School of Management and Economics Beijing Institute of Technology. Her specialty is data mining and technology innovation management.

References

Barney, J. 1991. “Firm Resource and Sustained Competitive Advantage.” *Journal of Management* 17 (1): 99–120.

Blei, D. M., and J. D. Lafferty. 2006. "Dynamic Topic Models." Proceedings of the 23rd International Conference on Machine Learning, 56–63. New York: ACM

Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. “Latent Dirichlet Allocation.” *Journal of Machine Learning Research* 3: 993–1022.

Breschi, S., F. Lissoni, and F. Malerba. 2003. “Knowledge-relatedness in Firm Technological Diversification.” *Research Policy* 32 (1): 69–87.

Brockhoff, K. K. 1992. “Instruments for Patent Data Analyses in Business Firms.” *Technovation* 12 (1): 41–59.

Brockhoff, K. K. 2002. "Indicators of Firm Patent Activities." Technology Management: The New International Language. IEEE.

Cantwell, J., and F. Fai. 1999. “Firms as the Source of Innovation and Growth: the Evolution of Technological Competence.” *Journal of Evolutionary Economics* 9 (3): 331–366.

Cho, I., and M. Park. 2015. “Technological-level Evaluation Using Patent Statistics: Model and Application in Mobile Communications.” *Cluster Computing* 18 (1): 259–268.

Cui Z, X. Zhao, X. Gao, D. Li, and W. Zuo. 2018. “Research on the Distribution Structure of Cereal Crop Seed Technology Innovation Based on Patent.” *Agricultural Science & Engineering in China* 2: 29–36.

Eck, N. J. V., and L. Waltman. 2010. “How to Normalize Cooccurrence Data? An Analysis of Some Well-Known Similarity Measures.” *Journal of the Association for Information Science & Technology* 60 (8): 1635–1651.

Ernst, H. 1998. “Patent Portfolios for Strategic R&D Planning.” *Journal of Engineering & Technology Management* 15 (4): 279–308.

Fai, F. M. 2005. "Using Intellectual Property Data to Analyse China's Growing Technological Capabilities." *World Patent Information* 27 (1): 49–61.

Garcia-Vega, M. 2006. "Does Technological Diversification Promote Innovation?: An Empirical Analysis for European Firms." *Research Policy* 35 (2): 230–246.

Griffiths, T. L. 2004. "Finding Scientific Topics." *Proc. National Academy of Sciences* 101 (1): 5228–5235.

Grindley, P. 2010. "Turning Technology into Competitive Advantage." *Business Strategy Review* 2 (1): 35–48.

Guan, P., Y. Wang, and Z. Fu. 2016. "Effect Analysis of Scientific Literature Topic Extraction Based on LDA Topic Model with Different Corpus." *Library & Information Service* 60 (2): 112–121.

Herrero, A., E. Corchado, L. Saiz, and A. Abraham. 2010. "Dipkip: A Connectionist Knowledge Management System to Identify Knowledge Deficits in Practical Cases." *Computational Intelligence* 26 (1): 26–56.

Hu, Z., F. Shu, and L. Tian. 2014. "Empirical Study of Constructing a Knowledge Organization System of Patent Documents Using Topic Modeling." *Scientometrics* 100 (3): 787–799.

Jian, C. G. 2006. "A Study of the Relationship Between Competitiveness and Technological Innovation Capability Based on DEA Models." *European Journal of Operational Research* 170 (3): 971–986.

Kajikawa, Y., K. Abe, and S. Noda. 2006. "Filling the Gap Between Researchers Studying Different Materials and Different Methods: a Proposal for Structured Keywords." *Journal of Information Science* 32 (6): 511–524.

Koellinger, P. 2008. "The Relationship Between Technology, Innovation, and Firm Performance – Empirical Evidence From e-Business in Europe." *Research Policy* 37 (8): 1317–1328.

Lai, K. K., and S. J. Wu. 2005. "Using the Patent co-Citation Approach to Establish a new Patent Classification System." *Information Processing & Management* 41 (2): 313–330.

Lee, W. S., E. J. Han, and S. Y. Sohn. 2015. "Predicting the Pattern of Technology Convergence Using big-Data Technology on Large-Scale Triadic Patents." *Technological Forecasting & Social Change* 100: 317–329.

Leonard-Barton, D. 2010. "Core Capabilities and Core Rigidities: A Paradox in Managing new Product Development." *Strategic Management Journal* 13 (S1): 111–125.

Ma, T., R. Li, G. Ou, et al. 2017. "Topic Based Research Competitiveness Evaluation." *Scientometrics* 2 (3): 1–15.

Marikkannan, M., M. Marikkannan, and A. Kannan. 2008. "An Intelligent System for Semantic Information Retrieval Information From Textual Web Documents." International Workshop on Computational Forensics. Springer-Verlag, 135–146.

Mayindi, D. H., and M. O. Kachienga. 2008. "Analysis of National Technological Competitiveness: South Africa's Civil Aircraft Industry." Portland International Conference on Management of Engineering & Technology. IEEE, 2008, 99–107.

Nakamura, H., S. Suzuki, I. Sakata, and Y. Kajikawa. 2015. "Knowledge Combination Modeling: The Measurement of Knowledge Similarity Between Different Technological Domains." *Technological Forecasting & Social Change* 94 (116): 187–201.

Narin, Francis, N. Elliot, and P. Ross. 1993. "Patents as Indicators of Corporate Technological Strength." *Research Policy* 16 (2): 143–155.

Patel, P., and K. Pavitt. 1997. "The Technological Competencies of the World's Largest Firms: Complex and Path-Dependent, but not Much Variety." *Research Policy* 26 (2): 141–156.

Pavitt, K. 1985. "Patent Statistics as Indicators of Innovative Activities: Possibilities and Problems." *Scientometrics* 7 (1–2): 77–99.

Porter, A. L., A. S. Cohen, J. D. Roessner, and M. Perreault. 2007. "Measuring Researcher Interdisciplinarity." *Scientometrics* 72 (1): 117–147.

Prahalad, C. K., and G. Hamel. 1990. "The Core Competence of the Organization." *Harvard Business Review* 68 (3): 79–91.

Ramani, S. V., and M. A. D. Looze. 2002. "Using Patent Statistics as Knowledge Base Indicators in the Biotechnology Sectors: An Application to France, Germany and the U.K." *Scientometrics* 54 (3): 319–346.

Suominen, A., H. Toivanen, and M. Seppanen. 2016. "Firms' Knowledge Profiles: Mapping Patent Data with Unsupervised Learning." *Technological Forecasting & Social Change* 115: 131–142.

Venugopalan, S., and V. Rai. 2015. "Topic Based Classification and Pattern Identification in Patents." *Technological Forecasting & Social Change* 94: 236–250.

Wang B, S. Liu, K. Ding, and Z. Liu. 2015. "Patent Content Analysis Method Based on LDA Topic Model." *Science Research Management* 3: 111–117.

Wang, M., H. Lo, and Y. Liao. 2015. "Knowledge Flow Determinants of Patent Value: Evidence From Taiwan and South Korea Biotechnology Patents." *International Journal of Innovation & Technology Management* 12 (3): 251–266.

Wernerfelt, B. 1984. "A Resource-Based View of the Firm." *Strategic Management Journal* 5 (2): 171–180.

Yau, C. K., A. Porter, N. Newman, and A. Suominen. 2014. "Clustering Scientific Documents with Topic Modeling." *Scientometrics* 100 (3): 767–786.

Yoon, J., and K. Kim. 2012. "Detecting Signals of new Technological Opportunities Using Semantic Patent Analysis and Outlier Detection." *Scientometrics* 90 (2): 445–461.

Yu, W. D., and S. S. Lo. 2009. "Patent Analysis-Based Fuzzy Inference System for Technological Strategy Planning." *Automation in Construction* 18 (6): 770–776.

Zhang, Y., Y. Guo, X. Wang, D. Zhu, and A. L. Porter. 2013. "A Hybrid Visualisation Model for Technology Roadmapping: Bibliometrics, Qualitative Methodology and Empirical Study." *Technology Analysis & Strategic Management* 25 (6): 707–724.