Patent-based technology forecasting: case of electric and hydrogen vehicle

Samira Ranaei* and Matti Karvonen
LUT School of Business and Management, Lappeenranta University of Technology, FI-53851 Lappeenranta, Finland
Email: samira.ranaei@lut.fi
Email: matti.karvonen@lut.fi
*Corresponding author

Arho Suominen
VTT Technical Research Center of Finland, FI-02044 VTT Espoo, Finland
Email: arho.suominen@vtt.fi

Tuomo Kässi
LUT School of Business and Management, Lappeenranta University of Technology, FI-53851 Lappeenranta, Finland
Email: tuomo.kassi@lut.fi

Abstract: The purpose of this paper is to study the evolution of emerging technological paths in the automotive industry. Worldwide patent data from the years 1990–2010 was collected and utilised to define the technological life cycles of the electric and fuel cell vehicle technologies. The novelty of our study is practicing the patent analysis approach using text mining techniques to collect patents according to their concepts in the automotive industry. The patent analysis results are compared to existing literature and expert opinion studies in alternative fuel vehicles field. The findings suggest that the development of electric vehicles will be quicker with a higher R&D share, compared to hydrogen vehicles. By gathering data and insights, the paper also offers general views on future automotive technology trajectories.

Keywords: technology forecasting; patent analysis; text mining; technology life cycle; electric vehicle; hydrogen vehicle; zero emission vehicle.


Biographical notes: Samira Ranaei works as a doctoral student in the School of Business and Management at Lappeenranta University of Technology. She received her Master’s in Innovation Management from Lappeenranta...
University of Technology in 2013. Her research interests are in innovation and technology forecasting, technology management and patent analysis techniques.

Matti Karvonen is a Post-doctoral Researcher in the School of Business and Management at Lappeenranta University of Technology. He received his master’s degree from Joensuu University in 1999 and Lappeenranta University of Technology in 2004, followed by the doctoral degree from Lappeenranta University of Technology in 2011. His research interests are in intellectual property management, industry evolution, convergence, and patent analysis methods.

Arho Suominen received his doctoral degree from University of Turku, Finland. Suominen had previously received his MSc and BSc from the University of Turku and also holds a degree from the National Defense University. His research interests include technology foresight, research policy and technology management. Currently, he holds the position of Senior Scientist at the Innovation and Knowledge Economy unit at the VTT Technical Research Centre of Finland, while also lecturing at the Department of Information Technology at the University of Turku.

Tuomo Kässi is a Professor Emeritus of the school of Business and Management at Lappeenranta University of Technology. He has an MSc in Technical Physics and in Economics. He wrote his doctoral thesis in Industrial Engineering and Management. He has worked in research and development, and sales and marketing; and as the director of machinery production and a consultant in industry some 25 years before entering his academic career. His research interests are in technology management, innovation and strategy.

This paper is a revised and expanded version of a paper entitled ‘Forecasting emerging technologies of low emission vehicle’ presented at PICMET Conference, Kanazawa, Japan, 27 July 2014.

1 Introduction

The transition towards sustainable transportation options is a response to increasing environmental regulations (Nieuwenhuis and Wells, 1997), the depletion of fossil fuels and climate change. So far the technological limitations, high costs and risks have kept the automotive industry focused on incremental innovations rather than propel it into a radical shift to alternative engine technologies. Currently, the internal combustion engine (ICE) and its more environmentally friendly version, hybrid electric vehicles (HEVs) are the dominant components of unsustainability in today’s car industry. Two competing solutions are primarily working towards a sustainable option, the electric vehicle (EV) and hydrogen vehicle (HV). These vehicles are called zero-emission vehicles (ZEVs), because they have no local exhaust pipe emissions. EV and HV contain a full or partial electric engine and subsystem technologies that are different from the ICE. Both technological variety and organisational competition in these alternative vehicle technologies have been increased over the last years (Oltra and Saint Jean, 2009b).

There are several challenges on the path towards ZEVs. There is a lack of infrastructure or alternative fuel production systems, known as the chicken or the egg
problem (Browne et al., 2012). The technological transition is a radical shift for the automotive industry, and the transition requires great investments (Utterback, 1996; Christensen, 1997). In addition, the growing awareness of environmental issues together with concerns about competitiveness have created considerable interest in research of advanced technologies for highly efficient ICE (Berggren and Magnusson, 2012; Oltra and Saint Jean, 2009a). Also, the development trend of EVs and HVs have been characterised as a hype, which was attributed to the high optimism period of their commercialisation followed by the disappointment of customers and stakeholders, specifically in the case of fuel cells (Suominen et al., 2011). It remains unclear what the dominant design of sustainable automotive transportation will be.

Recent research shows optimistic projections for the ZEV. The technologies are currently more expensive than conventional passenger cars, but projections suggest that they will become less costly thanks to technological learning (Weiss et al., 2012). The number of companies producing EV models has substantially increased (Sierzchula et al., 2012). HV development shows a promising future as well. Despite the ‘hydrogen hype’ (Bakker, 2010b), which caused huge disappointment, recent research (Rizzi et al., 2014) shows that the interest in hydrogen technologies in the automotive industry has not decreased over time.

The aim of this paper is to analyse the development of the ZEV through patenting. Even though patents are a mere proxy for actual development, patenting behaviour in the automotive sector has been shown to reflect actual research and development efforts (van den Hoed, 2005). Our proxy analysis with patents draws from connecting patent classifications to ZEV technologies and using the identified patent classifications as a source for the patent sample, subsequently used to model the developments of ZEVs. The methodological novelty of the approach is circumventing the caveats of using international patent classification (IPC) as a sole basis for sample creation by classifying patents to relevant and non-relevant patents using text mining in automotive industry. The results of the modelling will be compared with existing literature and expert opinion studies to enrich the interpretation.

The remainder of this paper is structured as follows. In the next section we provide the theoretical background to the patent analysis approach and the peculiarity of alternative technology vehicles in the automotive industry. The methodology section presents our patent search strategy on collecting relevant patent data, and afterwards plotting patents with the life cycle approach. The result and discussion will be presented in the last sections.

2 Background

The basis of a patent analysis lies in understanding the managerial implications of the results. The managerial implications of patent analysis have been drawn from multiple bodies of literature, such as technological evolution (Abernathy and Utterback, 1978) and the resource-based view of the firm (Barney, 1991; Chen, 2011). Patent analysis implicitly creates a window to quantify the knowledge embedded in organisations. This study draws from the literature on technology cycles (Abernathy and Utterback, 1978), looking at the competence-destroying transformation in the automotive industry and subsequent changes in the knowledge transformations embedded in organisations.
Technological evolution is composed of three main phases (Anderson and Tushman, 1991; Tushman, 1997): firstly, technological discontinuities trigger the periods of technological and competitive ferment with many technology options and a lot of market uncertainty among producers and customers. Secondly, in the course of the evolution of a technology, an era of standardisation emerges, during which product design and market needs stabilise and a dominant design emerges. Thirdly, the emergence of a dominant design opens an evolutionary path based on incremental innovation activities, focusing on refining the existing design. Technological discontinuities have been normally viewed as a threat to industry incumbents (Tushman, 1997) where technologies are emergent and more radical in nature. In these circumstances entrants enjoy potential competitive advantage over incumbents. The described pattern repeats itself when a new technology with the potential to substitute the older one is introduced. Patent information can quantify the patterns of technological cycles, the introduction of a discontinuity and the subsequent emergence of a dominant design.

Considering patents as the source of information to learn about technological development has both benefits and pitfalls. The major benefit of patent data is its uniqueness, meaning that the information stored in patents may not be republished in non-patent literature like books or articles (Liebesny, 1974), and its rich technological information is accessible online with a long time series. On the other hand, patent counts analysis per se does not reflect variation in technological quality or the commercialisation of inventions (Verspagen, 2007).

Patent data remains a unique resource for the study of technical change (Griliches, 1990). It represents a valuable source of technical information that can be used to quantify the evolution of technologies over time (Daim et al., 2006). Patents can be used to measure the impact of R&D activities (Ernst, 1997), diffusion of technologies and technological trajectories (Liu and Shyu, 1997). Patent analysis can support the study of prioritisation of R&D programs (Jeon et al., 2011) and technological assessment of competitors (Narin et al., 1987).

In the automotive industry, patent analysis has been used to understand the technological transformation of alternative vehicle technologies (Pilkington, 2004; van den Hoed, 2007; Oltra and Saint Jean, 2009b; Frenken et al., 2004; Yang et al., 2013). A majority of the studies have applied conventional patent analysis strategies, using keywords and ‘elementary measures’ (Suominen, 2013). Recent studies have developed from the mere quantification of countries, authors or technologies and augmented the analysis with text mining (Yau et al., 2014).

The assumption in patent analysis with the text mining approach is that the high probability of occurrence of one specific term or phrase (combination of a few words) in a title or abstract of patent documents is a better indicator for its relevance to one technology area or industry. Traditional patent search queries were formed, for instance, ‘electric’ AND ‘vehicle’ in an attempt to search for EVs. These types of queries produce patents that can be related to any type of vehicle that uses electric current (Wesseling et al., 2014). Therefore, the high number of irrelevant patents reduces the reliability of the study results. On the other hand, considering the patents that represented a high rate of occurrence of the phrase ‘electric’ with ‘brushless motor’ and ‘vehicle’ somewhat convey the contextual meaning of the document which may be related to EVs. Another dominant patent search strategy is to use patent classification codes (e.g., IPC). However, they are considered too broad to be directly applied to a technology area of interest.
In the automotive industry, the development of emission-free vehicles is in the state of ferment (Pohl and Yarime, 2012; Sierzchula et al., 2012), associated with a high level of uncertainty about the evolution of alternative technology vehicles (Contestabile et al., 2011). The uncertainty related to the technology evolution and prospects can be reduced by understanding the technology cycles. Indicators such as patents, publications or citations have been introduced as a major asset in technology forecasting (Watts and Porter, 1997). Patent applications, particularly in the manufacturing sector such as the automotive industry, can be effectively utilised to forecast and scrutinise the trends of technological activities (Yoon, 2012).

While the technology evolution model of Abernathy and Utterback presents the three main technology cycles, Watts and Porter (1997) suggested a method known as the technology life cycle (TLC) that can identify the technology development stages. The TLC approach suggests a linear five-step development model identified by bibliometric methods. The five cycles, basic research, applied research, development, application and social impact can be plotted by the number of patents, scientific publications or daily news databases. Despite the criticism over the linearity of the model (Rosenberg, 1994), it remains a practical illustration of technological life cycles (Balconi et al., 2010).

The TLC model is based on the assumption of a sigmoid growth curve. It assumes that following basic research, applied research saturates through a sigmoid growth pattern for which a patent application serves as a proxy measure (the number of patent applications over time generally follows an S-shaped curve) (Ernst, 1997). The curve starts with a slow progress, followed by an exponential growth after which the technology reaches saturation. The stage of saturation, or maturity, implies that the technology maybe is in the process of being substituted by a new technology or a new generation of an existing technology. Extrapolating patent data with time series on a given technology at a given time using the S-shaped curve as a reference links patent data to the technology cycle. This straightforward analysis is not without caveats related to the nature of patent data (Watts and Porter 1997) and challenges related to the limits of the S-shaped growth model (Suominen and Seppänen 2014). By understanding the limitations, however, patent data can create a powerful managerial tool.

3 Research method

The research design is based on creating a patent data proxy for the technology cycle of ZEV technology and modelling the technological development to the future. The technological lifecycles of the EV and HV will be plotted based on the number of patents to illustrate the evolution of these alternative technologies in the automotive sector. As illustrated in Figure 1, our research process includes

a) the collection of patent data from the European Patent Office (EPO) Worldwide Patent Statistical Database (PATSTAT, 2013), supplemented with previous literature and expert opinions

b) and c) the application of text mining and machine learning algorithms to filter out patents irrelevant to the EV and HV from our sample

d) the comparison of patent analysis results with expert opinion and existing literature on EV and HV development, to provide meaningful future projections.
The patent data is retrieved from the PATSTAT database and limited to a period from 1990 to 2010. The literature shows that the ZEV development intensified since 1990 (Kemp, 2005), which makes it a practical starting point. The years 2011 to 2013 are excluded due to the restrictions related to patent application becoming public. We did not use the patent family filter as emphasised by Wesseling et al. (2014) and Oltra and Saint Jean (2009a) as the same inventions registered in several countries shows their crucial value. Therefore, the high weight of one specific invention in our data set indicates its importance.

Pilkington et al. (2002) showed that analysis based on patent classification is of limited value in studying EVs. They investigated EV development by using the patent class search B60L11 resulting in a significant amount of irrelevant patents included in the sample. The IP classification includes a wide range of EVs, not just automobiles, and therefore the patents included within this classification relate to many other applications apart from EVs. Moreover, the ZEV consists of several components and emerging technologies that are not all categorised under one patent class and may be assigned in more than one class simultaneously.

The study by Pilkington et al. (2002) emphasises that a study should use clear boundaries between generic patents related to electric device technologies and automotive-oriented patents. This can be achieved by using an archive of reliable keywords (Rizzi et al., 2014; Frenken et al., 2004; Wesseling et al., 2014). However, the
problem of keywords is the inconsistency of how terminologies are used by companies, researchers or attorneys. In addition, the database search is based on the match of exact wording; it is merely a means of finding phrases without contextual meanings. Also, being unfamiliar to the technology area, for which the patent data are being gathered, it would be quite difficult to build an exhaustive keyword list.

To improve the accuracy of the patent retrieval process, we adopted the patent analysis method using machine learning and text mining previously practiced on semiconductor technology (Wu et al., 2010). The text-mining approach for patent analysis shown to be a practical technique in other technology areas (Tseng et al., 2007; Yoon and Park, 2004; Ranaei et al., 2014). The focus of these methods is on changing the unstructured part of patents (e.g., title, abstract), which contains valuable technical information, to structured data (numbers) to facilitate its analysis.

In this study RapidMiner software and its text-processing extension was used to classify documents and filter out irrelevant patent documents from the sample. The aim is to filter irrelevant patent documents, in respect to the EV and HV, from the large collection of patents listed under the vehicle section of IPC green inventory. The green inventory is a list of patent classification developed by IPC committee experts to help users to retrieve information on environmentally sound technologies (ETS). The IPC green inventory, while significant, is not an exhaustive list of patents, and additional IPC classes relating to ultra-capacitor types, hydrogen storages or electric motors were added based on expert opinion. As a result, we formed the query (see Appendix A) based on a broad IPC collection on PATSTAT, which resulted in a sample of 174,188 patent documents (named as our test set in Figure 1).

To select the relevant patents to our target technology, the supervised learning process used requires a training set validated by experts. The training set contains the most relevant phrases and the relevancy of test set keywords will be assessed against them. To collect patent for the training set, another set of queries was structured based on IPCs and keywords (see Table 1). These phrases and patent classes were collected from a literature review (Rizzi et al., 2014; Chan, 2007; Ball and Wietschel, 2009; Bakker, 2010a; Chen et al., 2011; Sierzchula et al., 2012) and expert interviews.

Table 1 Technology options and key components of EV and HV

<table>
<thead>
<tr>
<th>Vehicle technology alternatives</th>
<th>EVs</th>
<th>HVs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Propulsion motor</strong></td>
<td><strong>Battery types</strong></td>
<td><strong>Conversion type</strong></td>
</tr>
<tr>
<td>PM synchronous or PM brushless motors</td>
<td>Lead acid</td>
<td>ICE</td>
</tr>
<tr>
<td>Induction motor</td>
<td>Nickel base</td>
<td>PEM fuel cell</td>
</tr>
<tr>
<td>Switched reluctance motors (SRM)</td>
<td>Lithium base</td>
<td>Metal hydrides</td>
</tr>
<tr>
<td></td>
<td>Na-Nicim</td>
<td>On-board reforming</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Carbon material</td>
</tr>
</tbody>
</table>

Notes: aPermanent magnet

The sample retrieved for the training set was manually screened, reviewed and classified by experts. The process which took about five weeks, included several rounds of manual
screening and interviews with experts. Then followed by the actual manual classification of relevant HV and EV patents by help of six experts. As a result, the EV training set includes 147 relevant and non-relevant patents where the relevant patents comprise a balanced selection of patents relevant to the key technologies of electric motors (brushless motor, induction motor, permanent magnet, etc.) and battery types (lithium and nickel based, lead acid, etc.), and the non-relevant class contains all other patents not related to the EV, like the ICE and fuel cell patents. A similar categorisation process was done to structure the HV training set of 193 relevant and non-relevant patents.

After creating the training and test sets, we conducted the patent classification process in four phases (Steps b and c in Figure 1). Firstly, text pre-processing was applied on the training set to reduce its complexity. The pre-processing methods used included stop-word filtering and n-gram identification. The data was thereafter tokenised and transformed using the TF-IDF\(^2\) scheme. This process yields a matrix that reflects the keyword occurrence probability measure (Appendix B). Secondly, a support vector machine (SVM) classifier (Cortes and Vapnik, 1995) was selected to learn the TF-IDF scores of all true and false categories. For instance, considering the EV training set, the TF-IDF score of ‘combustion engine’ will be zero in the relevant category but higher in the non-relevant category. Therefore, the SVM classifier will label a patent document as irrelevant, if it indicates a high TF-IDF score for the phrase ‘combustion engine’.

Thirdly, a cross-validation process was conducted using the training set to show the classification accuracy of the system. The classifier indicated 90.45% and 95% accuracy levels for the EV and HV, respectively. Fourthly, the text-processing step was replicated for the test set. The trained SVM was then applied to the TF-IDF matrix of our test set and the patents were classified into two categories. The final results labelled as relevant patent documents to the EV (16,595 docs) and HV (10,291 docs) were plotted on the growth curves. An overlap of about 0.04% was identified between the two categories, which is acceptable since we had small training sets and 5% classification errors.

Good forecasts usually draw upon a range of perspectives and methods (Porter et al., 2011). To achieve comprehensive interpretations we utilised patent analysis, growth curves, expert opinion and reviewing the literature on hydrogen and EV development (Step d in Figure 1). Regarding expert opinion, we kept in contact with the experts during the structuring of our patent classification system and used their comments for our result interpretation. Furthermore, the comparison between the patenting behaviour of the two green vehicle technologies was conducted based on previous literature on prototype and production analysis (Sierzchula et al., 2012), prototype and basic research (Sjoerd et al., 2012) and patents (Wesseling et al., 2014).

4 Results

In this section we present the main findings related to the patent trends, assignees contribution, analysis of new entrants per year and industry diversity of applicants.

4.1 Trend analysis: competition within and between trajectories

The relevant patent documents contained two main alternative technology categories, EVs and HVs using either the PEM\(^3\) fuel cell or ICE as an energy conversion unit. The
number of retrieved patents on the EV and HV, were 16,595 and 10,291, respectively, during a time period of 1990–2010.

Figure 2 The number of patent applications of EV and HV 1990–2010

Distribution by a patent publication year is an indicator that quantifies the intensity of R&D within a technology area. Based on Figure 2, it can be argued that both technology options have received increased attention after 1990, partly owing to the concerns over development of green technologies. The patent dynamic between the two alternative technology vehicles, confirms that EVs are attracting a larger share of green-oriented R&D in the automobile industry, also supported by the findings of Sierzchula et al. (2012).

Figure 3 Normalised number of patent applications of EV and HV 1990–2010

The periods of optimism and disappointment known as hype periods for hydrogen cars and EVs have affected the patenting behaviour. The changes in the patenting behaviour are more visible in the unit-normalised diagram, where both trends are illustrated on the same scale between zero to one (see Figure 3). What is evident from the unit-normalised number of patent activities is the three phases of attraction shift between the two vehicle alternatives. Previous studies of the EV and HV, based on other data sources like prototypes and magazine articles, illustrated a similar competition trend (Sjoerd et al., 2012).
Following the innovation activities of two technology alternatives illustrated by previous diagrams, Figures 4 and 5 represent the proportion of existing patent applicants, rate of new entrants and their contributions to the development of HV and EV over the years. In this paper, the number new entrants are defined as new applicants in each year that have not been appeared in previous years (Suominen, 2013). To simplify the applicant analysis on our large data sets, individual assignees and those firms with less than five patents are excluded. Figure 4 exhibits higher number of industry players and new entrants for EV comparing to HV. Similarly in Figure 5, the proportion of EV patents from both existing applicants and new comers exceeds its competitor’s share of contribution.

Figure 6 is demonstrating industry diversity of patent applicants. Both EV and HV assignees have been clustered in seven groups based on their industry origin. Since the patenting activity before 21st century was relatively low for green technology vehicles, only the years between 2000 and 2010 are considered. The highest proportions of EV applicants are from automotive companies, electronic industry and conglomerate firms. While in case of HV, besides automotive industry, applicants from fuel cell systems and other industries (biotechnology, mechatronics, oil and gas) have highest contributions.

Also, it is worth to note that share of universities and research institutes applied for EV
related patents have slightly increased by later years. Moreover, rate of new entrants signals the attraction of specific technology area and the level of market dynamic. It can be observed from Figure 7, that the number of HV new entrant has been considerably declined after 2007, while EV is gradually attracting rising number of new industry players.

**Figure 6** Industry diversity of (a) EV and (b) HV assignees 2000–2010 (see online version for colours)

To provide a clearer picture of the competition between HV and EV development in this paper, detail comparison of patenting behaviour is explained in three periods of time. During the first decade starting from 1990 to 2000, EVs were dominating the automotive industry (Figure 3). The number of EV prototypes between 1990 and 1997 is reported much higher than its technology competitor: 47 EV vs. nine HV prototypes (Sjoerd et al., 2012). During this time car producers were concerned with the low range between recharging and finding a suitable battery type. The high number of EV patenting or prototypes indicate an era of experimentation and exploration. Additionally, the time lag between prototypes (applied research) and patent (development) perhaps suggests the applicability of a technology life cycle model in the automotive industry. The EV trend shows even higher patenting activity by 2000, which is three years after the announced EV prototypes.
Figure 7  Industry diversity of new entrants for (a) EV and (b) HV 2000–2010 (see online version for colours)

The second phase of competition from 2001 to 2007, illustrates a higher share of HV related patents. Figure 6 also presents a dramatic increase in HV patent applicants in this period. Furthermore, Figure 7 depicts the rate of new entrants in HV market which is at its highest for three years in a row (2002–2004) before a dramatic decline occurred by 2007. In addition, prior studies on car prototypes have reported 33 number of HV prototypes between the years 1998–2005, and only nine EV prototypes (Sjoerd et al., 2012).

Third phase started from mid-2007, an upward surge in the patenting activity of EVs can be observed (Figure 3). Unlike HV, number of EV assignees and even new entrants show a steady growth toward 2010, which may signal the return of EV with high potential of commercialisation (Figures 6 and 7). Previous literature argued that the most important driving force behind EV returning to the picture are; increasing number of car manufacturers, appearance of start ups and convergence of car companies toward a single mutual battery type (Sierzchula et al., 2012).

4.2 Future development paths for electric and hydrogen cars

The extrapolation of EV and HV patent trends provide perspectives on the technology life cycle of these technologies and estimate future R&D developments (see Figure 8).
The patent data was modelled using the Fisher-Pry function and fitted data with a high R² coefficient of 0.99. The estimated value of a, b and L of Fisher-Pry function are reported on Appendix C. Although Figure 3 suggests that EVs are moving forward with a higher proportion of patents than HV, it seems that EV will have a considerably higher share of light-duty vehicle fleet in the future than today.

**Figure 8** Trend extrapolation of the EV and HV

The HV growth curve currently shows development on the technology life cycle and will enter the maturity phase and later the plateau level approximately by 2016. Although the graph shows the saturation point for hydrogen cars, we cannot directly interpret that there will be no further advancement. The development of hydrogen cars hinges on the key technology components and various types of subsystems. Considering the PEM fuel cell as one of the key technology subsystems, the R&D activity may shift from a low-temperature PEM fuel cell to a high-temperature PEM fuel cell which would lead to a new S-curve (Mock and Schmid, 2009). On the other hand, technological development seems insufficient to explain fuel cell success in the automotive industry, and other vital issues like environmental regulations should be considered (van den Hoed, 2007).

In the case of EVs, it can be observed from the short forecast that the saturation phase will eventually appear, but within a longer time lag. Currently, the wave of EV development is being reinforced by intense competition, the high rate of new entrants and industry diversity of the involved market players. Also, as the experts commented, the possibility of sharing the present infrastructure and technologies established for hybrid cars may encourage more investments in the EV, rather than the HV. The experts also added that the adoption of the EV may differ in terms of geographical difference. For example, in Finland the present electric heating platforms established for winter time in parking lots can serve electric cars for the recharging purpose. Therefore, in the case of the transition to EVs, countries such as Finland may have the advantage of less initial investment in infrastructure and a more promising future with EV adoption. The predicted EV growth until 2016, or even after this, will continue if the influential factors and situation remain unchanged. All in all, EV appeared to have a considerably higher share patenting and technology development activities than HV and so the market expectations of EV light-duty vehicles seem to be higher than HV.
5 Discussion

The use of patent data is gaining increased interest in the field of technology management and technological forecasting. Patents represent technological evolution and include speculation about how that particular technology might be developed and used in the future. To gain insight into the development trend of the green vehicle alternative, historical and patent analyses were performed. By contrast to previous patent studies in the automotive industry, we applied text mining and machine learning approaches to retrieve and classify the patent documents, which resulted in a more comprehensive technology development study. Furthermore, extrapolation methods were applied to provide information on the future trajectory of emission free cars. In order to provide a meaningful interpretation and reliable forecast, expert opinion, previous literature and prototype analysis of electric and hydrogen cars were integrated into our results.

The reasons behind utilizing text mining and machine learning in patent search and data retrieval in this paper are the following:

1. Patent classes are too wide to analyse a specific technology field, and their hierarchical structure is based on technology fields rather than an application area.
2. New emerging technologies are not assigned to a specific patent class.
3. The inconsistency between the keywords entered by researchers, innovators and applicants may return irrelevant patents.
4. The keyword search strategy may not capture the contextual meaning.

In response to these shortcomings of conventional patent search strategies, the application of a patent classification system yields more accurate patent data more efficiently. Once the training and test sets are defined, performing patent classification on free open source software only takes a few clicks on the mouse. The classification accuracy is measurable; in case of undesired results the training sets can be revised and updated under expert supervision. The possibility of modifying the presented classification process reflects its flexibility.

Significant growth for both the EV and HV, mainly triggered by the California Air Resources Board (CARB) regulation from 1990, was illustrated by patent trend analysis. A clear shift of industry interest between the two options during three distinct time periods was confirmed by integrating previous literature results, prototype analysis and vehicle production information (Sierzchula et al., 2012; Sjoerd et al., 2012; Wesseling et al. 2014). In the first phase, EV development took a larger share of R&D activities as from 1990 for a decade. Due to the technical and market limitations for EV development, a window of opportunity opened for the HV and attention gradually turned toward fuel cell vehicles by 2000. Finally, after seven years, by early 2008 electric cars returned, reinforced by a higher number of industry actors and new entrants.

Trend extrapolation was conducted to give insight into the future development of the two major green vehicle options. The S-shaped growth models have projected continuing growth for the EV and an early saturation level for hydrogen cars by 2016. Although extrapolations are powerful forecasting tools, the interpretations need to be structured with caution. We have learnt that several forces influence the development of electric cars (Wesseling et al., 2014), but extrapolation does not consider those factors. In fact, extrapolation methods lack casualty (Suominen and Seppänen, 2014), which underlines
the need for complementary methods, revision and updating the forecast result by comparing it to actual development. This is specifically true in the case of modelling a technology such as fuel cells, which has been under significant policy interventions (Suominen, 2014). The saturation of HV patenting growth may even suggest a need for a discontinuity, a new generation of fuel cells, developed through a radical innovation that could push the technology to a new growth trajectory. As it stands now, the rapid saturation of HV technology suggests a commercialisation path that may be unfeasible.

For EVs the technology life cycle suggests a path more in line with actual developments. According to the recent report of US Department of Energy (DOE), the sale of EV (BMW Active, Ford Focus, Honda Fit, Tesla Model S, RAV4, Chevrolet Spark, Smart ED and Mitsubishi I) has been dramatically increased from 10,111 cars in 2011 to 23,112 by 2013 (USDOE, 2015).

Overall, the comparison made over patenting trends with other sources, shows that electric cars development may move forward with higher speed compared to HVs. Using an improved approach of patent data collection and supplementing the results with other sources of information assumed to provide more accurate support to the decision-making process. The extrapolation of patents to the future has its limitations, and too straightforward managerial implications can be misguided. As seen with the HV, the trend path seems unfeasible and without the understanding of how the hype around the technology impacts the model, we are bound to incorrect inferences.

The exploration of our patent analysis method is so far limited to the automotive industry field. This area appears suitable for text mining and classification since the keyword trends appeared simple. The method may yet not be ready to be generalised for sophisticated patent fields, like chemistry with complex symbols and chemical compounds. The language barrier is another limitation of this study, since only English patents have been considered. Future work can be directed to extract more information from patent text, such as application of unsupervised learners to identify the topical shift among pile of patent documents.

6 Conclusions

Owing to the rising demand for green technology in the automotive sector and uncertainty over the adoption of a future car technology, we have explored the technological development of hydrogen and electric vehicles. This study, firstly, utilised the automated patent analysis approach to retrieve relevant patents with application in the car industry. As a result, the patent search and analysis was accomplished more quickly with a higher accuracy in the data retrieval process.

Secondly, the competition between hydrogen and electric cars has been illustrated based on patent data and supplemented with various information sources including expert opinion and previous literature. Results indicated

a significant growth for both technology alternatives since 2000
b continuing development in future but with different speed.

However, in terms of R&D proportion, the EV has a greater share compared to hydrogen cars. Overall findings moreover suggest a faster and higher rate of development for the EV, if all external forces remain unchanged.
Although the forecasting results with extrapolation tools provide only rudimentary information, they can still shed some light on the direction of the technological trajectory. Decisions should be made, even though enough information is not available. Researchers and patent analysts can benefit from the cost-effective and quicker patent retrieval process since we used free open source software.

References


Notes
1 Listed by WIPO: http://www.wipo.int/classifications/ipc/en/est/
2 Term frequency –Inverse document Frequency
3 Proton Exchange Membrane
4 \[ y(t) = L/(1 + ae^{–bt}) \]

Appendix A

### IPC classes used in the test set query

<table>
<thead>
<tr>
<th>IPC green inventory</th>
<th>MOTOR TECHNOLOGY</th>
<th>BATTERY TECHNOLOGY, BATTERY MANAGEMENT, STORAGE OPTIONS</th>
<th>ELECTRIC VEHICLE CONTROLLING SYSTEM</th>
<th>ADDITIONAL IPCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brushless motors</td>
<td>Electric propulsion with power supply from force of nature, e.g., sun, wind</td>
<td>Electric propulsion with power supply external to vehicle</td>
<td>• With power supply from fuel cells, e.g., for hydrogen vehicles</td>
<td>• Hybrid propulsion systems comprising electrical and internal combustion motors, and storage of electric energy</td>
</tr>
<tr>
<td></td>
<td>Electric propulsion with power supply from force of nature, e.g., sun, wind</td>
<td>Electric propulsion with power supply external to vehicle</td>
<td>• Combustion engines operating on gaseous fuels, e.g., hydrogen</td>
<td>• Arrangements in electrical propulsion in connection with power supply from force of nature, e.g., sun, wind</td>
</tr>
<tr>
<td></td>
<td>Electric propulsion with power supply from force of nature, e.g., sun, wind</td>
<td>Electric propulsion with power supply external to vehicle</td>
<td>With power supply from fuel cells, e.g., for hydrogen vehicles</td>
<td>Electric equipment or propulsion of electrically-propelled vehicles</td>
</tr>
<tr>
<td></td>
<td>Electric propulsion with power supply from force of nature, e.g., sun, wind</td>
<td>Electric propulsion with power supply external to vehicle</td>
<td>Combustion engines operating on gaseous fuels, e.g., hydrogen</td>
<td>Control systems of fuel cell</td>
</tr>
<tr>
<td></td>
<td>Electric propulsion with power supply from force of nature, e.g., sun, wind</td>
<td>Electric propulsion with power supply external to vehicle</td>
<td>Power supply from force of nature, e.g., sun, wind</td>
<td>Methods of moulding or forming tanks or container for hydrogen storage</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>B60L 9/00</th>
<th>B60L 11/18</th>
<th>F02B 43/00</th>
<th>B60K 16/00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H02K 29/08</td>
<td>B60L 8/00</td>
<td>B60L 9/00</td>
<td>B60K 17/00</td>
</tr>
<tr>
<td></td>
<td>H01M4/13, H01M10/0525, H01M8/04, H01M8/0, H01M8/24, H01M2/02, H01M 16/00, H01M 4/00</td>
<td>B60K 8/00</td>
<td>B60K 8/00</td>
<td>H02K 17/00</td>
</tr>
<tr>
<td></td>
<td>H02K 29/08, H02k 17/00</td>
<td>B60K 1/04, B60K 8/00</td>
<td>B60K 11/02</td>
<td>B60K 15/00, B60K 15/03</td>
</tr>
<tr>
<td></td>
<td>B60L 8/00, B60L 9/00, B60L 9/18, B60L 7/14</td>
<td>B60L 8/00, B60L 9/00, B60L 9/18, B60L 7/14</td>
<td>B60L 8/00, B60L 9/00, B60L 9/18, B60L 7/14</td>
<td>B60L 8/00, B60L 9/00, B60L 9/18, B60L 7/14</td>
</tr>
<tr>
<td></td>
<td>B60w 10/26, B60w 10/28, F02M 21/06</td>
<td>B60w 10/26, B60w 10/28, F02M 21/06</td>
<td>B60w 10/26, B60w 10/28, F02M 21/06</td>
<td>B60w 10/26, B60w 10/28, F02M 21/06</td>
</tr>
<tr>
<td></td>
<td>B29C 63/08, B29C 53/08</td>
<td>B29C 63/08, B29C 53/08</td>
<td>B29C 63/08, B29C 53/08</td>
<td>B29C 63/08, B29C 53/08</td>
</tr>
</tbody>
</table>
**IPC classes used in the test set query (continued)**

- Vehicle body, substructures and sub-units: B62D 21/18, B62D 25/20
- Electric or fluid circuits or arrangements of elements thereof specially adapted for vehicles, batteries and carrying-off electric charges, radiators: B60R19/00, B60R 16/02
- Rider propulsion of wheeled vehicles: B62M 6/90
- Testing of vehicles, engines: G01M 16, 17/00
- Cooling systems for vehicle, controlling of coolant flow: F01P 7/12
- Hydrogen; Gaseous mixtures containing hydrogen; Separation of hydrogen from mixtures containing it: C01B 3/00

### Appendix B

**TF-IDF matrix for the electric vehicle training set (partial view)**

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Total occurrence</th>
<th>Doc occurrence</th>
<th>EV</th>
<th>Not_EV</th>
</tr>
</thead>
<tbody>
<tr>
<td>battery</td>
<td>1078.0</td>
<td>122.0</td>
<td>903.0</td>
<td>175.0</td>
</tr>
<tr>
<td>electric</td>
<td>732.0</td>
<td>138.0</td>
<td>557.0</td>
<td>175.0</td>
</tr>
<tr>
<td>lithium</td>
<td>657.0</td>
<td>49.0</td>
<td>657.0</td>
<td>0.0</td>
</tr>
<tr>
<td>lithium_ion_battery</td>
<td>511.0</td>
<td>36.0</td>
<td>511.0</td>
<td>0.0</td>
</tr>
<tr>
<td>vehicle</td>
<td>501.0</td>
<td>136.0</td>
<td>118.0</td>
<td>383.0</td>
</tr>
<tr>
<td>motor</td>
<td>415.0</td>
<td>87.0</td>
<td>251.0</td>
<td>164.0</td>
</tr>
<tr>
<td>automobile</td>
<td>385.0</td>
<td>60.0</td>
<td>366.0</td>
<td>19.0</td>
</tr>
<tr>
<td>charging</td>
<td>354.0</td>
<td>68.0</td>
<td>276.0</td>
<td>78.0</td>
</tr>
<tr>
<td>electric automobile</td>
<td>311.0</td>
<td>55.0</td>
<td>305.0</td>
<td>6.0</td>
</tr>
<tr>
<td>combustion</td>
<td>297.0</td>
<td>35.0</td>
<td>2.0</td>
<td>295.0</td>
</tr>
<tr>
<td>control</td>
<td>296.0</td>
<td>94.0</td>
<td>114.0</td>
<td>182.0</td>
</tr>
<tr>
<td>device</td>
<td>296.0</td>
<td>81.0</td>
<td>167.0</td>
<td>129.0</td>
</tr>
<tr>
<td>ignition</td>
<td>243.0</td>
<td>25.0</td>
<td>0.0</td>
<td>243.0</td>
</tr>
<tr>
<td>fuel_cell</td>
<td>232.0</td>
<td>51.0</td>
<td>21.0</td>
<td>211.0</td>
</tr>
<tr>
<td>permanent magnet</td>
<td>208.0</td>
<td>30.0</td>
<td>208.0</td>
<td>0.0</td>
</tr>
<tr>
<td>magnet_synchronous</td>
<td>201.0</td>
<td>30.0</td>
<td>201.0</td>
<td>0.0</td>
</tr>
<tr>
<td>permanent_magnet_synchronous</td>
<td>201.0</td>
<td>30.0</td>
<td>201.0</td>
<td>0.0</td>
</tr>
<tr>
<td>internal_combustion</td>
<td>194.0</td>
<td>33.0</td>
<td>0.0</td>
<td>194.0</td>
</tr>
</tbody>
</table>

Note: Explanation: it can be easily observed that, for example, the TF-IDF score of ‘internal combustion’ in the last row is 0 in the electric vehicle category and 194 in the irrelevant to electric cars category. A similar pattern is visible for permanent_magnet_synchronous, lithium-ion_battery, etc. Some neutral or generic keywords like device, control or battery that are being used in a variety of applications have more or less occurred in both categories.
Appendix C

Estimated parameters for the Fisher-Pry function

<table>
<thead>
<tr>
<th>Technology cluster</th>
<th>L</th>
<th>a</th>
<th>b</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electric vehicle</td>
<td>27.0156</td>
<td>160.6386</td>
<td>0.2775</td>
<td>0.9998</td>
</tr>
<tr>
<td>Hydrogen vehicle</td>
<td>11.5341</td>
<td>927.7430</td>
<td>0.4499</td>
<td>0.9997</td>
</tr>
</tbody>
</table>