Technological Forecasting & Social Change 115 (2017) 37–51

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Forecasting by analogy using the web search traffic

Seung-Pyo Jun *, Tae-Eung Sung, Hyun-Woo Park

Department of Industrial and Market Analysis, Korea Institute of Science and Technology Information, 66, Hoegi-ro, Dongdaemun-gun, Seoul 130-741, Republic of Korea

ARTICLE INFO

Article history:
Received 19 March 2016
Received in revised form 7 July 2016
Accepted 15 September 2016
Available online 26 September 2016

Keywords:
Analogical forecasting
Web search traffic
Product life cycle
Search cycle
Google trends
Naver trend

ABSTRACT

Various types of demand forecasting methods have been developed and utilized to predict the adoption of new technologies. Recently, along with the advancement of bibliometrics, there have been particularly active attempts to forecast life cycles using technology documents such as news, paper publications, patents, etc. The present study uses web search traffic to forecast by analogy, which has newly emerged as a method of empirically verifying the life cycle of either a product or a technology. So as to explore the potential of the analogical forecasting method using search traffic, we compare the trends of changes in the life cycle with those of search traffic and compare aspects of the search traffic exhibited by both U.S. and Korean consumers over various products. The study results revealed that search traffic trends tended to precede the adoption of a new product; however it accounted for the trends of adoption over the full life-cycle very accurately. In addition, statistically significant relationships have been observed in the search traffic for the same technology even when the traffic originated from distinct nations, languages and web search engines. From the results therein, we judged that the search traffic-based, analogical forecast method would be effective, and applied it to a case for estimating the Korean Plug-in Hybrid Electric Vehicle (PHEV) market. The most significant contribution of this study is that it presents the potential of utilizing search traffic as a new dimension for forecasting by analogy.

1. Introduction

In order to identify or assess a research field with converging technologies or trends, a variety of methodologies in Webometrics have been developed and suggested to date (Jabłońska-Sabuka et al., 2014; Vaughan and Yang, 2013). In addition, in terms of quantitative analysis, the concepts of product life cycle (PLC) or technology life cycle (TLC) are already widely and frequently used in various academic fields including business management, technology management and science and technology policy development. Recently, along with the advancement of Webometrics, there have been particularly active attempts to analyze life cycles through a quantitative analytical approach and to utilize the results in forecasting (Daim et al., 2006; Wei et al., 2012). This study investigates the possibility of connecting the analogical forecasting method to web search traffic information, which is an instance of Big Data accumulated from the traces of consumer behavior, and thereby presents the potential of using this web search traffic to achieve a more objective analysis of the life cycle and the adoption of a new technology. This study seeks to contribute to Webometrics or relevant interdisciplinary fields by exploring the method of utilizing search traffic, a type of Big Data, for the purpose of forecasting innovative products and analyzing their life cycles.

When seeking to analyze sales or change of a specific market, if the sales size or change of the existing market is known, i.e., when there are many types of data that are directly available, it is possible to utilize numerous quantitative estimations in forecasting. However, in cases where there is of scarcity in information regarding a specific market such as in the case of a new technology product, we are able to analogize the market by making use of information on other similar technologies. In such cases, when comparing two markets or technologies, similarity and importance are critical for conducting a structured analogy (Martino, 1993). The study herein pays attention to the availability of search traffic as a dimension (or attribute) of comparison in analogy forecasting. Although analogy estimation (or analogy forecasting) is a representative technique of qualitative analysis, search traffic is a typical example of quantitative data, and we expected that bringing these two together would have a strong synergistic effect. Market sales (size) information, in which we are mainly interested when forecasting, is difficult to measure, requiring much time and expense. By contrast, search traffic is relatively easy to measure, objective, and inexpensive to measure. Therefore, if we utilize the existing market information, which has been so laboriously obtained, as well as the recent search traffic data, this would render analogy forecasting more widely applicable. Technological or economic dimensions that have been hitherto used for analogical forecasting are useful indicators, but collecting such data

* Corresponding author.
E-mail addresses: spjun@kisti.re.kr, spjunn@msn.com, spjun@ust.ac.kr (S.-P. Jun), ts322@kisti.re.kr (T.-E. Sung), hpark@kisti.re.kr (H.-W. Park).

http://dx.doi.org/10.1016/j.techfore.2016.09.014
0040-1625/

for comparison demands much effort; for this reason, search traffic, which has the advantage of being analyzable in real time based on big data, can serve as a highly useful proxy variable (or dimension).

So as to evaluate the feasibility of this approach, we must first examine whether search traffic is capable of strongly explaining demand. Also, we must verify whether search traffic is significant as a comparative indicator in markets that are different from one another. We need confirmation in regards to the two prerequisites for any dimension of analogy forecasting, namely importance and similarity. Accordingly, this study firstly investigates whether search traffic accurately predicts market sales or change of new technology, i.e. its life cycle. Next, the study compares search traffic data for the same new technology in disparate markets. Lastly, by making full use of forecasting by analogy using search traffic, we attempt to make an estimate for a new market, an area in which forecasting has hitherto remained difficult. Using a variety of information regarding the U. S. market and search traffic information for both the U.S. and Korea, we predict a new PHEV (Plug-in Hybrid Electric Vehicle) market in Korea.

In order to describe the contents and results of our study, we begin in Section 2 by reviewing preceding research related to the forecasting methodology by analogy, utilized in both the consumer adoption model and the consumer behavior model, which provide the foundational theories for this study overall. We then cite preceding studies related to consumer search behavior and web search traffic information to clarify the significance of utilizing web search traffic, which is the focal point of this study. In Section 3, we explain how our research methodology is uniquely differentiated from other conventional methodologies, and then proceed to explain the key sources of data used and the method of data collection. Also, to validate whether our approach is meaningful in the field of forecasting, we present several selections of case studies and analyze their implications. In Section 4, we compare the trends of changes in search traffic with the aspects of market penetration for products that have just entered the decline phase in their product life cycles while conducting comparisons across the respective countries and markets as well as analyzing the worldwide market. We then compare search traffic in the United States and Korea for major IT products (new technologies). In Section 5, In order to validate both the accuracy and superiority of the methodology proposed, we compare prediction performance of three analogy methods and six mathematical prediction models regarding practical cases with the overall sales achievement already known. In order to illuminate the usability of this study, we also apply analogy forecasting to PHEV sales in Korea, using both search traffic and sales by developed countries.

This study is expected to ultimately contribute to a better understanding and theorization of the life cycle using webometrics and will also improve our ability to utilize analogical forecasting. The search traffic analyzed in this study is a type of data that will prove highly useful for observing and forecasting consumer behavior with greater speed and economy.

2. Theoretical background and preceding studies

For successfully predicting the adoption or sales of a product that applies new technology, we must understand the consumers who adopt those technologies. Consumers undergo various processes when adopting new technology. Marketing theory defines adoption as the decision of an individual to become a regular user of a certain product (Kotler and Keller, 2008). In innovation theory, adoption is regarded as the most advanced step of the process, since it represents the decision to utilize the innovation without any reservation (Rogers, 2003). This section briefly reviews the concept of analogy estimation (or forecasting). Next we review the key theories developed regarding adoption, and examine theories related to consumer behavior. Lastly, we proceed to introduce preceding studies that have examined information searching behavior and web search information, which are the key elements of the adoption process on which we focus in this study.

2.1. Analogical forecasting

It must be recognized at the outset that forecasting by analogy is essentially a qualitative method. It cannot produce numbers. Nevertheless, it is a conscious and deliberate attempt to draw upon historical experience. Forecasting by analogy involves a systematic comparison of a technology to be forecast with some earlier technology that is believed to have been similar in all or most important respects. But what does it mean to be "similar," and which respects are "important"? Answering these questions is the whole point behind the idea of systematic comparisons (Martino, 1993).

SAS (SAS, 2012), a widely-used business analytics software provider company, mentions various situations for new product forecasting, i.e. entirely new types of products, new markets for existing products, etc., and emphasizes structured analogy as one of the primary approaches in new product forecasting, such as top management opinion, Delphi method and anonymous group interviews. Using the structured analogy approach, Green and Armstrong (Green and Armstrong, 2007) described a judgmental procedure that uses information from conflict decision making situations in a structured way.

Martino (Martino, 1993) observed that there are several issues that can emerge when utilizing analogies and argued for the need to address these problems. These problems are: lack of inherent necessity, historical uniqueness, historically conditioned awareness, and casual analogies. Firstly, lack of inherent necessity simply means that the outcome of a historical situation is not completely determined by physical factors. The fall of an object, for instance, is completely determined by factors such as gravity, air resistance, initial velocity, and so on. Knowing these factors, we can "forecast" very accurately what the trajectory of the object will be. There are no similar "laws" governing the outcome of historical situations. Even though we know what happened, we find it hard to believe that people acted as they did. Analogies are based on the assumption that there is a "normal" way for people to behave and that given similar situations, they will act in similar ways. However, there is no guarantee that people today will act as people did in the model situation. Hence, the forecast is at most probable, never certain (Martino, 1993).

Secondly, historical uniqueness simply means that no two historical situations are exactly identical in all respects. Thus, it is important to be able to say which comparisons are important and which can be ignored. An analogy will be strengthened if there are several historical cases with compatible outcomes that can be compared to the present case to be forecast. However, since each of these historical cases is unique in itself, it is important to be able to determine whether they are "similar enough" to each other to be considered analogous. Thus, a systematic means for comparing model situations with each other and with the current situation is essential.

Thirdly, the problem of historically conditioned awareness arises because people may be aware of what happened at "the last time." Even though a historical situation may be judged to be sufficiently similar to the present situation to be called analogous, people may be aware of the previous outcome. If they do not like the way the previous situation turned they may deliberately act differently this time in order to secure a more preferred outcome. Historically conditioned awareness, then, violates the assumption that there is a "normal" way for people to behave and that they always behave that way. Thus, despite a forecaster's best efforts to check for analogous situations, the forecast may be invalidated by people's awareness of prior outcomes (Martino, 1993).

Lastly, a casual analogy arises when we observe that two things are alike in a few aspects, and assume without evidence that they are alike in most other aspects. It is possible, of course, that such an analogy is valid. However, in most cases it will be in error. It is not sufficient to observe that two things are alike on one or two characteristics. It is necessary to make a systematic check of all the observable characteristics before you can risk concluding that the two are alike on as of yet unobservable characteristics. Despite these problems, analogies can be a very useful method for forecasting technological change. The problems cannot be solved completely, but they can be minimized by using a systematic method for establishing analogies (Martino, 1993). To resolve the limitations discussed above, Martino (Martino, 1993) and Thomas (Thomas, 1985) advise us to systematically compare various dimensions, and identifies technological, economic, managerial, political, social, cultural, intellectual, religious-ethical, and ecological dimensions as dimensions that must be taken into account.

In regards to the limitations and advantages of analogical forecasting, Goodwin and Wright (Goodwin and Wright, 2010) claimed that it performs relatively strongly in predicting cases where there is a sparsity in the reference class but is unable to offer strong predictions where there is an inappropriate reference class or misplaced causality. This leads us to conclude that analogical forecasting can be a useful prediction method for the adoption of new technology where there is hardly any relevant data available, but caution must be exercised when selecting data for comparison or when comparing dimensions. Taking into consideration the selection of data for comparison and comparisons among the dimensions, Goodwin et al. (Goodwin et al., 2012) compared the results of numerous analogy methods, by explaining how to apply to analogy of the existing similar products as an alternative while explaining the limitations of the new product analysis on the utilization of mathematical models (Bass model) which are widely and often used in demand forecasting. Authors compared the predictive power of a number of American consumer electronics products by four methods of published values, random selection of analogies, nearest neighbor analysis, and regression analysis. According to the study results, Goodwin et al. (Goodwin et al., 2012) presented that prediction throughout most analogy methods as a whole reveals significant errors, and the averaged analogy method based on the parameters of multiple products improves the accuracy more than a single product's analogy result itself. However, if we increase similar products up to five to six or more in applying to the Bass model, there is little difference in accuracy; and when similarity with new products are weighted and considered, it revealed that analogy results in multiple products return the higher accuracy than a single product case. Recently, Goodwin et al. (Goodwin et al., 2014) critically approached the preceding studies regarding the three methods of management judgement, analysis of judgements by potential customers and formal models of the diffusion process which are often used in demand forecasting of durable products newly launched. Authors pointed out risks and uncertainty the limitations of prediction by a single method would have, and presented that formal models of the diffusion process are slightly higher in accuracy than unstructured judgement, where analogy helps improve the accuracy of formal models. However, they identified that analogy with multiple products as much is essential to enhance the accuracy, and well-structured the issues to consider in selection and application of earlier products in analogy method while discussing the characteristics of products such as length of time that the comparable must hold.

Analogical forecasting is often referred to as ‘historical analogy’ or ‘life cycle analogy,’ and is sometimes used in linkage with various other methods. A leading example of this is the use of analogical forecasting in predictions based on diffusion models such as the Bass model. Since the publication of Bass models in 1969, Mahajan et al. (Mahajan et al., 1990) predicted by analogy the integrational coefficient of influence ($p + q$) based on external and internal parameter values, i.e. the coefficients of innovation (p) and imitation (q). Earlier on, Bayus (Bayus, 1987) demonstrated that demand forecasting can be made by analogizing the life cycles of products which are highly correlated such as software (consumables or accessory items subject to their repeat purchase), and hardware (the original durable goods required for use of their software). Analogical forecasting is often used in regards to the diffusion of new technologies as well as the adoption of products, and it can be used to analogize the differences in the time of introduction across various countries. Grübler et al. (Grübler et al., 1999) claimed that the diffusion pattern of new technology appears to be similar to those of existing technologies. They introduced the concept of co-evolution in explaining the diffusion of energy technologies and demonstrated through empirical analysis based on far-reaching, historical data that the technology and infrastructure of technology followers and also the diffusion process of their energy sources follow those of developed countries or technology leaders in a closely similar manner. Analogical forecasting is also one of the principal prediction methodologies in bibliometrics. Daim et al. (Daim et al., 2006) presented the forecasts for three emerging technology areas by integrating the use of bibliometrics and patent analysis into well-known technology forecasting tools such as scenario planning, growth curves and analogies.

Analogical forecasting has been in frequent usage in more recent research as well. Lee et al. (Lee et al., 2012) performed demand prediction using an analogy-based diffusion model, which applied the concept of the life cycle. For the purpose of forecasting the broadband internet services market, they simultaneously utilized the life cycle of an existing product and the data on consumer reservation prices collected through surveys. Also, Routley et al. (Routley et al., 2013) argued that although the life cycle is frequently used for analogy, it is ill-defined and often transposed, and proposed that using technology roadmapping (TRM) architecture can be an effective alternative for explaining the development of a technology or industry. The concept of analogical forecasting has been recently applied to determine a reasonable R&D investment policy and technology penetration rate in renewable energy for Korea, where Kim et al. (Kim et al., 2014) considered the renewable energy diffusion process based on the relationship between the diffusion pattern and R&D investment drawn by analogy from an empirical case available from the market of Germany, an advanced country. Kim et al. (Kim et al., 2014) then determined how investment triggered the growth of technology in Germany and applied the S-curve relation formula to derive an appropriate investment in Korea.

In regards to the use of analogical forecasting methodology, this study follows the existing methodology in that it makes use of the comparison of a similar technology's life cycle analogy and the comparison between technology leaders and followers. However, there exists an obvious distinction in that the research herein provides search traffic as data in order to compare the differences among life cycles or countries.

2.2. Consumer adoption model and consumer behavior model

In technology innovation theory, the consumer adoption model is the key theoretical component for explaining consumer behavior. In this context, innovation pertains to goods, services or ideas that people perceive as new, regardless of how long they have existed. The above definition of innovation is based on the work of Rogers (Rogers, 2003), who defined the innovation diffusion process as the “the dissemination of a new idea generated by invention or creation among the final users or adopters.” In other words, the consumer adoption process scrutinizes the mental experiences that the individual undergoes from the moment he or she first hears of the innovation up to the stage of final adoption. The adopters of new products proceed through five stages in the adoption process, namely the stages of Awareness, Interest, Evaluation, Trial and Adoption (Kotler and Keller, 2008). The approach to the adoption process outlined by Rogers highlights the mental process experienced by the individual. It therefore qualifies as an approach from the point of view of the user (or consumer), in contrast to the life cycle, which is a conventional approach that focuses on the producers. Among the five stages listed above, the Interest stage is the one most closely related to search traffic, which is the subject of this study.

Rogers (Rogers, 2003), in the course of explaining the differences in the preparation taken for the usage of new products and the impact of individual factors, defined the innovativeness of an individual as the relative speed of a particular individual in adopting new ideas compared to other members within the social system. In each product field, there are innovators, early adopters, early majority, etc. On the basis of these classifications of adopters and their characteristics, some researchers have argued that there are discontinuities in adoption. The argument is that there is a discontinuity in adoption, or "chasm," between the "innovators and early adopters" and the "early majority" (Moore, 1999). There have also been studies that used data from the consumer electronic industry to empirically demonstrate that there are instances where a sales peak in the early-staged market is followed by a slump, which is then followed by growth into another peak exceeding the initial market peak. This phenomenon has been referred to as a "saddle" (Goldenberg et al., 2002; Peres et al., 2010). The "saddle" is similar to the concept of a "chasm" in that it emphasizes the discontinuity between the adoption behavior of early adopters and main market adopters, and its pattern is very similar to the hype. In terms of technology adoption by consumers, Hirunyawipada and Paswan (Hirunyawipada and Paswan, 2006) presented consumer innovativeness trait as a useful predictor of new product adoption, by validating the positive effect in both statements that cognitive, needs-specific innovativeness enhances the actual adoption of new products and that social and physical risks enhance consumer's propensity to information search on new products, on the contrary to financial risk.

The consumer behavior model also provides an important theoretical groundwork for the study of consumers' web information searches. The study of consumer behavior refers to the study of the methods used by individuals, groups and organizations to select, purchase, use and process products, services, ideas or experiences for the purpose of satisfying their primary and secondary needs. Such consumer behavior is influenced by cultural, social and personal factors, and numerous types of consumer behavior models have been proposed to better understand such consumer behavior. According to the stimulus-response model of buyer behavior, first, marketing stimuli and environmental stimuli enter the consciousness of the consumer (Kotler and Keller, 2008). The psychological process works in combination with special consumer characteristics to influence the decision making process and ultimately induce the decision to purchase. In other words, the consumer behavior model addresses the process that occurs within the consumer's consciousness from the time of the input of the external marketing stimuli to the time of the final buying decision. Associated with those consumer adoption and behavior patterns, Fernández-Durán (Fernández-Durán, 2014) analyzed the examples of products such as iPods, DVD players, and Wii Play video games to verify that almost all products exhibit seasonality in their sales patterns and seasonal effects can be influential in forecasting by analogy the weekly/monthly/quarterly sales of a new product using the seasonal Bass forecasting diffusion model.

When the decision making process for purchases is examined based on this consumer behavior model, it can be broadly distinguished into five stages, consisting of 1) problem awareness, 2) information search, 3) evaluation of alternatives, 4) decision to purchase, and 5) actions following the purchase. This model emphasizes that the process leading to the purchase is initiated long before the actual act of purchase, and that its results linger long after the purchase. However, consumers do not undergo all five stages in all cases when they purchase a product. In other words, some consumers sometimes simply skip over a certain stage, and some even move through some of the stages in inverse sequence. As in the consumer behavior model, the differences in cultural, social and personal factors inevitably result in diversifying the forms of adoption exhibited by consumers, particularly in their consumption of new products (Kotler and Keller, 2008). There are two implications here that must be particularly noted in relation to the present study. The first is that search traffic usually occurs in the information search stage, which is the second stage within the model. Secondly, there is the possibility that consumer behavior may vary depending on the country, language, or the specific society in question.

Meanwhile, researchers working from the perspective of behavioral science regard information searching by consumers to be involved in a stage of the purchase decision-making process and offer explanations of its cognitive aspects. In the Engel-Blackwell-Miniard model, which establishes the series of stages that constitute the purchase decision-making process of consumers, information searching is one of the activities that have an important influence on the purchases made by consumers (Engel et al., 1995). Moe (Moe, 2003), who researched the online information search activities of consumers, categorized these information search activities into browsing and searching according to the motivation and the outcome of the activities. Browsing is defined as the act of wandering around to read or view informational clues without any particular purpose, in contrast to searching, which is an action that takes place with intentionality for the purpose of accomplishing the task of purchasing or preparing for future purchases.

Since the development and nearly ubiquitous presence of the Internet and information technology have dramatically increased accessibility to sources of information and multiplied the efficiency and ease of acquiring information, consumers are now able to obtain information from a variety of sources through their web search activities. Through search engines, users can view many search results for the keyword that interests them and access a wide range of information sources. Given numerous search results, consumers do not utilize all of these available information sources and instead seek out specific sources to acquire the information they need. In other words, information searching by consumers involves decisions regarding which sources of information to select and use (Zimmer et al., 2007). Sproles et al. (Sproles et al., 1980) have found that consumers exposed to a larger quantity of information tend to have a higher level satisfaction regarding their final selection, and interpreted that more efficient consumers are those who tend to search for and utilize a larger amount of information. Also, the fact that consumers are searching for more information is understood to indicate that they are collecting information from a variety of information sources (Choo et al., 1999). On the other hand, Moon (Moon, 2004) analyzed three factors of people, problems and the context to enable consumers to adapt the Internet instead of traditional channel (such as direct visit to shops) for information search and product purchase. The hypotheses are 'Factors such as product type with pre-study or pre-experience of use before purchase, quality of products, price, delivery time-to-consumer and after-service directly influence the preference of consumer purchase through the Internet.', 'Factors such as consumer knowledge level, objective of information search, and demographic characteristics (gender, age, income, occupation) influence higher probability of employing the Internet for information search.', etc.

As noted in all of the above theories, consumers search for information in the process of adopting a new technology and purchasing a new product. In the following, we examine changes in the traffic of web searches, which is one of the leading means used by consumers for such information searches, for the purpose of analyzing expectations regarding innovative products and the characteristics of adoption.

2.3. Consumer search behavior and web search traffic information

When purchasing products or services, consumers search for relevant information in order to reduce financial and psychological uncertainties, avoid risks involved in the purchase, and ultimately maximize their satisfaction with a product. Due to the widespread availability of the Internet and the recent rapid dissemination of smartphones, the majority of consumers now use the Internet as the medium of their information searches. Also, the consumer's intention to search for information online is a key factor that affects the consumer's intention to purchase a product (Shim et al., 2001). The study by To et al. (P.-L. To et al., 2007) also demonstrated that there is a positive correlation between the degree of intention to search for information online and the degree of intention to purchase the product through online shopping malls. Kumar, et al. (Kumar et al., 2005) investigated how user-developed web search strategies could affect overall consumer search performance on the internet, and found out that consumer search behavior would be explained better in terms of search engine-based web search, which would provide a more complete view of users' actual search behavior.

The factors that can affect consumers' information searching activities can be broadly divided into the features of the product itself and the characteristics of the consumer who is conducting the information search. In regards to consumers who search for information, the consumers' knowledge has been treated as an important variable that influences the process of determining the intention to purchase, especially during the information search (Raju et al., 1995). It has been observed that an increase in the consumer's knowledge is accompanied by a corresponding increase in information search activities, but once the knowledge reaches a certain level, the consumer is no longer in need of additional searching and therefore the number of searches decreases. In other words, there is an inverse U relationship whereby a consumer with an intermediate level of knowledge conducts information searches more often than consumers with a high or low level of knowledge (Bettman and Park, 1980; Rao and Sieben, 1992). When purchasing products, a consumer's information search activity may vary depending on the type of product. Beatty and Smith indicated that more information search activity takes place in the cases where there is a higher perception of risk regarding the product (Beatty and Smith, 1987).

The web search activities of consumers can be analyzed through the web search traffic information provided by Google Trends. Many scholars have recently conducted a variety of studies on web search traffic information, and these studies can be broadly categorized into two distinct fields. The first field consists of studies directed toward making forecasts based on web search traffic information. Ginsberg et al. (Ginsberg et al., 2009) have developed a model that indicates the current level of the flu based on Google search traffic information. Although Butler (Butler, 2013) has recently challenged the reliability of the flu trend service offered by Ginsberg et al. (Ginsberg et al., 2009) using Google, this model is capable of providing flu information one or two weeks in advance of the report publicized by the Centers for Disease Prevention and Control (CDC), and is therefore currently considered to be useful (Ginsberg et al., 2009). Meanwhile, Choi and Varian (Choi and Varian, 2009) have shown that search traffic information can be used to forecast current economic activities such as car sales, house sales, retail and travel activities, etc. Goel et al. (Goel et al., 2010) have claimed that web search traffic can forecast rankings of feature films, first-month sales of video games, and the rankings of songs on the Billboard Hot 100 chart. Vosen and Schmidt (Vosen and Schmidt, 2011) have compared web search traffic information with consumption trend indices, specifically the University of Michigan Consumer Sentiment Index (MCSI) and the Consumer Confidence Index, and argued that web search traffic information provides a better forecast of consumer expenditures than such trend indices. Vaughan and Yang (Vaughan and Yang, 2013) have found significant correlations between web traffic and academic (or business) performance. Jun. et al. (Jun et al., 2014a) even demonstrates that search traffic pertaining to hybrid cars and their related attributes offers better explanations of consumer demand than conventional bibliometric analysis variables such as patents or news. Such preceding studies pertaining to search traffic show that search traffic can serve as a proxy measurement of social phenomena and can yield analytical results that are comparable to conventional surveys in providing macroscopic forecasts such as forecasts of demand and changes in consumption. In other earlier literature, Jun (Jun, 2012a) analyzed the hype cycles of three actors that constitute the core of the socio-technical system for the successful market entry of hybrid cars, where the hype cycle of the user is specifically analyzed by web search traffic.

As can be seen, although preceding studies have enhanced the applicability of search traffic, their limitation was that the studies were mostly restricted to a single site, namely Google, a single country, and one specific technology (product). On the other hand, this study takes a fresh approach to addressing the differences that exist across countries or search engines for the same technology, for the purpose of utilizing search traffic in analogical forecasting.

3. Research methodology and case studies

In this section, we first introduce the methodology used in this study and explain the research method used to collect the key data. Lastly, the section introduces the case studies selected for this study and explains the reasons behind their selection.

3.1. Research methodology

Recently, in response to the growth of the web, the concept of webometrics arose to refer to attempts to measure and evaluate all types of information available on the web. Webometrics is thus a new field of study that applies bibliometric methods to the web. In this sense, webometrics can be regarded as a methodology included under the category of bibliometrics (Thelwall, 2008; Thelwall et al., 2005). According to this explanation, we can say that the majority of the preceding studies cited above, which used news, paper publications, and patents, apply bibliometrics as their research methodology (Daim et al., 2006; Ruef and Markard, 2010; van Lente et al., 2013), and the studies that analyze information circulating on the web, such as search traffic, adopt the method of webometrics (Choi and Varian, 2009; Ginsberg et al., 2009; Jun et al., 2014a; Jun et al., 2014b).

In the present study, we collected data on the search traffic regarding specific technologies or products from the web to compare consumers' expectations with their actual adoption or sales pattern. More specifically, the search traffic data was collected from Google and Naver, which are respectively the dominant service providers for the market in the United States and Korea. In the case of Korea, since Korean is the language usually used by consumers, we chose to analyze cases in which the query was made using a search term in Korean rather than English. Specifically, we collected the search traffic for searches performed in Korean in cases where Korean was the more commonly used language for searches, in comparison to English (in the case of Blu-ray and Bluetooth, we compared the traffic for searches conducted in Korean).

Google's search statistics analyze a portion of web searches to calculate the number of searches for the terms input by a user within a specific time period in relation to the total number of searches conducted on Google. This is equivalent to expressing the probability that a particular individual user will search for a certain search term within a specific time period in a particular region. The search statistics set the criteria of minimum traffic for the search term and hence search terms with low search volume are not indicated in the statistics. Also, search terms that were repeatedly input by a particular user over a short time period is also excluded from the tally, preventing the possibility of artificially manipulating the level of interest through repetition (Google_Trends, 2014; Jun, 2012a). Naver also provides relative statistical values as its search traffic data, and the concept and process applied are identical to those of Google.

Since Google provides traffic information for each search term by country, it was possible to collect data specific to the United States or other individual countries (Google, 2016). Meanwhile, Naver only distinguishes between traffic from PCs and traffic from mobile devices (Naver, 2016), and for this study, we used the data on traffic from PC searches. According to Jun and Park (2016), analysis reveals that information search by PCs relatively well explains a product's sales, but since the study showed that the search records on the mobile devices did not commensurate with the sales achievement of a product yet, traffic information from mobile devices was excluded. Naver can be accessed from all over the world as there is no restriction on overseas access and search is available in both Korean and English, but the service is mainly utilized in Korea and most of information search is conducted in Korean (Naver_Trend, 2014). In this study, it is assumed that search traffic is reflective of national and linguistic differences. When we compare the two search engines on the basis of the description above, they have in common that both have their own predominant search services in the United States and Korea, respectively, and the language used for search query may be either English or Korean on both. The study focuses on the differences between the two engines. As for Google, the dominant language in the U.S. market is English. The engine also provides the search results by regions. Meanwhile, Naver is the primary search engine in Korea, where the dominant language is Korean. Although Naver has the limitation that it does not provide search results differentiated by region as Google does, we concluded that the lack of information control by geographical region on Naver is not a crucial issue, given that few people outside Korea speak and write Korean.

Both Google and Naver services allow users to download all search traffic results in a table format as well as a CSV file, and we chose to use the CSV file format. As for the market sales data used in this study, since there is no institution that provides a comprehensive service for this data, we relied on data provided by relevant institutions specific to each technology (product).

First of all, to secure more objective data on sales and life cycle, we compared special cases in which the life cycles of the items were almost at an end. Furthermore, we observed the time series features of the search traffic for each technology/product in a graph format and then compared the search traffics patterns across countries to examine the similarities and differences of consumer search behavior by country. To achieve greater objectivity in our analysis of the relation between the search traffic data and product sales, we additionally employed a statistical analytic method, using VECM (Vector Error Correlation Model) and Granger causality test to determine the causality, lead/lag structures and the similarities in search traffic.

3.2. Collection of data for analysis

Since we adopted webometrics as our research method, it was critical to use data that had been objectively collected and analyzed. For this reason, this study uses information provided by highly credible institutions or information made available by dominant corporations, as indicated in Table 1. Also, to ensure objectivity in data collection and to enable automated or systematic collection and utilization in the future, we relied on public information accessible via the web for all our data.

We measured the traffic of searches made on websites to quantify the expectations of users. For our search traffic analysis, we used the search statistics of Google, the U.S. search site that has a market dominance of over 66% (comScore, 2013) in U.S., as well as the search statistics of Naver, the dominant search site in Korea, which was used for the purpose of cross-country comparison. Naver's share of the search market in Korea is nearly 80% (Koreanclick, 2012). Although Koreans also use Google, the percentage is still very low, and the usefulness of the data is limited by the fact that only specific age groups are prone to use Google (Koreanclick, 2012). For this reason, we selected a local Korean search engine. The search traffic on Google or Naver was adopted as the index of consumer behavior for the reason that their search engines already occupy a monopolistic position in the market. Moreover, though producers also use Google (or Naver) searches, the majority of the Google (or Naver) users in this regard consist of consumers who are restricted from access to other specialized databases (DBs).

The advantage of utilizing Google trends is found in its process of normalization. Research case studies in the past used absolute values (for example, the number of hits, etc.) and hence failed to exclude environmental factors involved in consumer exposure, attributable to the overall increase in news volume or the number of web pages. By contrast, all of the results of the search statistics in the Google Trends data undergo a normalization process, dividing them by a common variable to eliminate the influence of variables. Google Trends data does not report the raw level of queries for a given search term. Rather, it reports a query index. The query index starts with the query share: the total query volume for a search term in a given geographic region is divided by the total number of queries in that region at a point in time. The query share numbers are then normalized so that they start at 0 on January 1, 2004. Numbers at later dates indicated the percentage deviation from the query share on January 1, 2004 (Choi and Varian, 2009). Naver benchmarked Google trend services and performed the normalization of traffic data as Google did. It presents the search traffic data based on search frequency and marks relative indicators out of the maximum 100 search query over the search period by aggregating the frequency of searches each week like Google. However, Naver does not explain normalization method in details as much as Google does (Naver_Trend, 2014). Through this method, it becomes possible to compare the basic features of each set of data. The market sales volume is also divided by the total sales volume to yield the sales market share of new products that is submitted to analysis.

3.3. Case studies

First, we compared search traffic to the life cycle of sales, a type of comparison that could not be performed in the preceding studies of the search traffic and sales pattern. Because it has been a relatively short time since search traffic data became publically accessible, we cannot observe the entire life cycle. To overcome this limitation, we sought out products that were released in the mid or late 2000s and have reached the end of their sales potential, and thus selected Netbook and Wii as our case studies. Wii is a home video game console released by Nintendo on November 19, 2006. As a seventh-generation console, Wii competes with Microsoft's Xbox 360 and Sony's PlayStation. Nintendo states that its console targets a broader demographic than that of the two others (Graves et al., 2010; Wikipidia, 2014). Netbooks are a category of small, lightweight, legacy-free, and inexpensive computers. At their inception in late 2007, as smaller notebooks optimized for low weight and low cost, the portable device referred to as Netbook omitted certain features (e.g., the optical drive), featured smaller screens and keyboards, and offered reduced computing power when compared to a full-sized laptop (Hodgson, 2008). Netbooks vary in size (generally from around 5″ to around 10″) but in features, generally have no CD-ROM or DVD drive, and fewer ports than a laptop. They have longer battery life (around eight hours) but have a smaller capacity (Pratt, 2010).

Secondly, to conduct comparisons of search traffic across products, we targeted IT products with relatively short life that exhibit change clearly. Moreover, to facilitate comparisons to preceding studies, we selected items that have already been the subject of preceding studies, including Blu-ray, Bluetooth and VoIP. Also, we additionally included IPTV, a product that was not covered in any preceding studies, to analyze the search traffic for innovation products in the United States using Google and in Korea using Naver. The first, Bluetooth, is an industrial specification for wireless personal area networks. It is a technology that provides a way to connect and exchange information between devices such as mobile phones and laptops. The name of the technology Bluetooth was adopted officially in 1998 (Järvenpää and Mäkinen,

Table 1
Data sources for each major variable.

Variables	Site or Sources	Explanation
Search traffic (U.S.)	Google trends	Weekly and monthly search traffic in the United States (2004-Present)
Search traffic (KOR)	Naver trend	Weekly and monthly search traffic in South Korea (2007-Present)
Netbook sales	iSuppli, IDC	Quarterly Sales (units) of Netbook (2007–2013)
Wii sales	Nintendo	Quarterly Sales (units) of Wii (2007–2014)
HEV & PHEV sales (U.S.)	Hybridcars.com	Monthly market share rate of new HEV and PHEV among new cars in the U.S. (2007-Present)

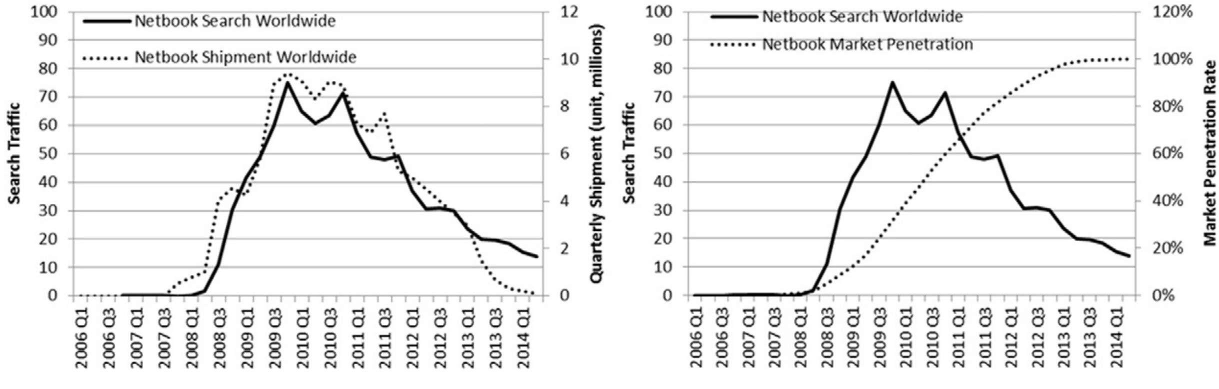


Fig. 1. Comparison of the worldwide search traffic for Netbooks to their quarterly shipment (left) and their rate of penetration (right).
Data source: Google (Google, 2016), Dignan (Dignan, n.d.) and Case (Case, 2012).

2008). The second technology Blu-ray is a high density disc format developed by Blu-ray Disc Association and it has won the battle to be the next generation optical storage disc. Blu-ray was officially announced in February 2002 (Fox, 2002) although the prototypes using the same blue-laser based technology were already unveiled at Ceatec (Combined Exhibition of Advanced TEChnologies) in 2000 (Järvenpää and Mäkinen, 2008). Internet telephony or Voice over IP (VoIP) emerged during the mid-1990s as a very promising technology that offered free long distance telephone calls. The first applications comprised a computer linked to the Internet and enabled the user to call another user with a computer and a VoIP software (van Lente et al., 2013; Wikipidia, 2014). IPTV (Internet Protocol Television) is a service that uses high-speed internet to deliver information services, video contents and broadcasts to TV. This service began in England in 1999 and expanded to the European and East Asian markets in the course of 2002 and 2003 (Wikipidia, 2014).

4. Research results

In this section, we target the special cases which allow us to compare the full life cycle to search traffic since this type of comparison has not been easy to perform in preceding studies. Next, we seek to identify specific characteristics of the search traffic by comparing search traffic for new IT technology, taking into account not only the differences in the products, but also the distinctions among countries, languages, and search engines.

4.1. Comparison of the search traffic and life cycle

Fig. 1 shows a comparison of the product life cycle of Netbook and the search traffic they generated. For the Netbook's life cycle, we based our analysis on the data on its sales volume (shipment) in the worldwide market up to 2012 and on the anticipated sales from 2013 onwards (Dignan, n.d.; Case, 2012). The search traffic data is based on the searches performed worldwide using Google. The graph on the left of Fig. 1 compares the quarterly shipment and search traffic, and the graph on the right of Fig. 1 compares the search traffic to the rate of penetration, determined on the basis of the accumulated shipment up to 2014. As seen in Fig. 1, Netbook has nearly reached a phase of decline in its life cycle. Also, the search traffic reached a peak at the time when the rate of penetration in the worldwide market was around 40%. In particular, we also observed that the search traffic for Netbook did show some correlation to the shipment of Netbook and the peak of search traffic preceded the growth stage of life cycle.

Next, Fig. 2 also shows a comparison of the search traffic for Wii and its product life cycle. The Wii's life cycle was analyzed based on its sales (hardware shipment) in the worldwide market up to the second quarter of 2014 (Nintendo, 2014), and the search traffic comprises searches conducted worldwide via Google. The left graph in Fig. 2 compares the annual shipment and the search traffic, while the right graph in Fig. 2 compares the search traffic to the rate of penetration, determined on the basis of the cumulative shipment up to 2014. As shown in Fig. 2, there was hardly any increase in the sales of Wii since late 2013, indicating that it has neared a phase of extinction. The comparison in Fig. 2 shows the peak of search traffic in the fourth quarter of 2008, while the peak of quarterly sales was in the fourth quarter of 2009, around a year later. The rate of penetration in the worldwide market was also around 40% when the search traffic reached a peak.

For some products, however, when comparing the two cycles at the level of the worldwide market, we need to take account of how factors such as the time at which the product was launched in each country and differences in purchasing power can influence the difference in the cycles. This complication is the reason why we limited our below case studies to specific countries when attempting to make comparisons

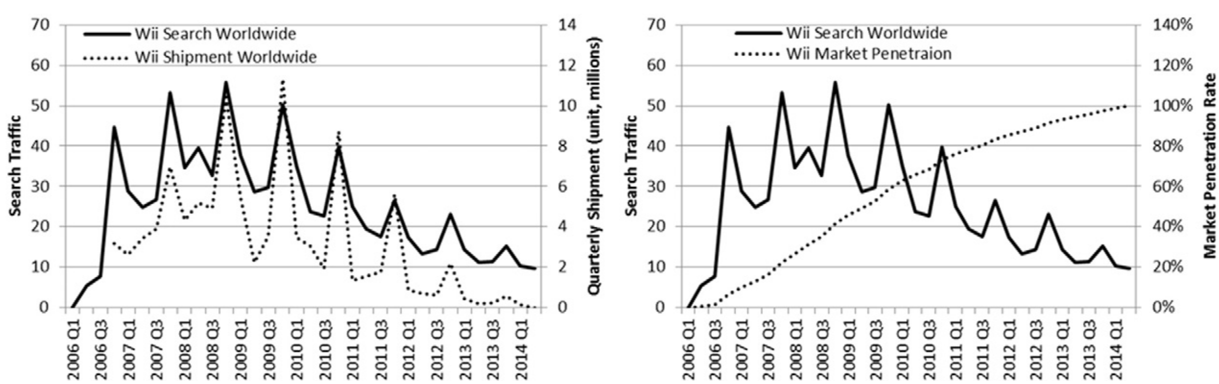


Fig. 2. The worldwide search traffic for Wii compared to its quarterly shipment (left) and rate of penetration (right).
Data source: Google (Google, 2016) and Nintendo (Nintendo, 2014).

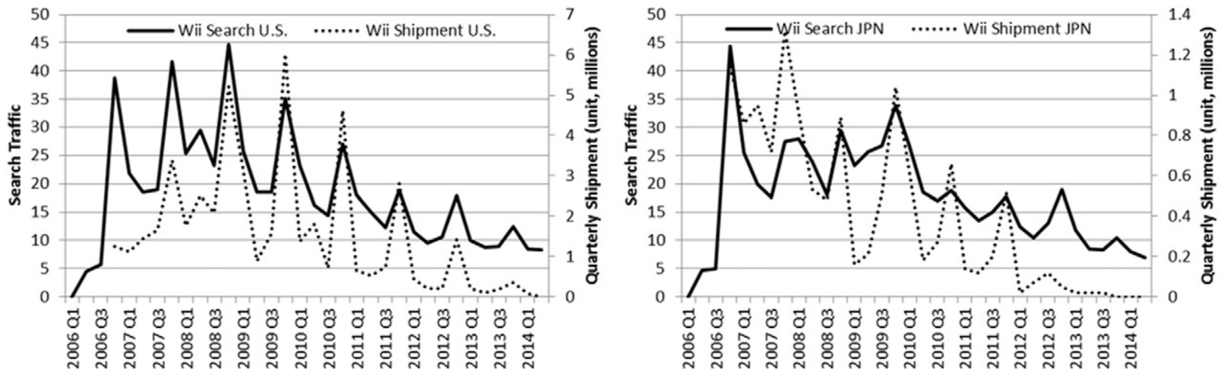


Fig. 3. Comparison of the quarterly shipment and search traffic of Wii in the United States (left) and Japan (right).
Data source: Google (Google, 2016) and Nintendo (Nintendo, 2014).

between countries or between products. Likewise, we limited our analysis of the life cycle of Wii and its search traffic to specific countries to take account of this complication. Fig. 3 analyzes the life cycle of Wii based on the actual quarterly penetration of Wii shipped into the markets in Japan and the United States up to the second quarter of 2014. For the search traffic as well, we limited our comparison to the Google search traffic occurring in each country. The market for Wii showed hardly any growth in Japan since the latter half of 2011 and in the United States since the second half of 2012, showing a faster decline relative to the decline in the worldwide market. Fig. 3 shows that the peak of search traffic occurred in the U.S. market around the time the rate of penetration reached 40%, while in the Japanese market, the peak of search traffic occurred when the rate of penetration was not yet 20%. The analytical data for each country shows that the quarterly sales was highest in the United States market in the fourth quarter of 2009, but the search traffic was highest in December of 2008. In the Japanese market, shipment was highest in the fourth quarter of 2007, while the search traffic was highest in December 2006. In both cases, the peak of search traffic preceded the peak in sales by around a year.

The above results, which compare the search traffic to the full life cycle, confirm that the peak of the search traffic precedes the peak in sales by a significant margin of time. On the scale of the worldwide market, the peak of search traffic is found to occur when the rate of penetration is around 20% to 40%. To verify the correlations between search traffic and sales, we performed the VECM Granger Causality/Block Exogeneity Wald Tests (hereinafter referred to as GCBEW tests). Before examining causality to test our time series data, we performed unit root tests to verify that the variables were stationary. We applied the Augmented Dickey-Fuller (ADF) and Dickey-FullerGLS (ERS) tests, which are commonly used in stationary tests (Mackinnon, 1996), and the results of these unit root tests are presented in Table 2 and Table 3.

First, Table 2 shows the unit root test results for the sales of each product. Unit roots were found to be present in the level variable for the majority of products, based on the results from at least one of the test methods. Therefore, the sales time series of the products are nonstationary and it would require a first or second order difference to make the series stationary; thus using a difference variable to determine significance in regards to sales would improve the accuracy of our findings.

Next, Table 3 presents the unit root test results for the search traffic of each product. In the level variable, we found a couple of cases in which unit roots were present in the results from one of the test methods with the exception of Wii (Japan). Therefore, for the search traffic time series of the products, we found it necessary to use a variable only in the case of Wii (Japan) with a second order difference to make the series stationary.

In the unit root test results discussed above, sales were observed to be nonstationary for the majority of products, and therefore we decided to use a difference variable to examine causality based on the VAR (Vector Autoregressive Regression) model. Prior to this, however, we needed to examine whether there is a long-run stationary relation between search traffic and sales. Studies by Granger (Granger, 1988) and by Bahmani-Oskooee and Alse (Bahmani-Oskooee and Alse, 1993) have shown that when unit roots are present in the data and the two variables are in a long-run stationary relation, it may be invalid to infer causality based on the usual Granger Test using a difference variable. These studies demonstrated that the results have stronger validity if the causality test is performed by using an error correction model that takes into account the long run balance between the variables. Therefore, we found it necessary to perform cointegration tests as a second stage, to determine whether the two variables have a long-run stationary relation. Cointegration test methods used in preceding studies include the methods proposed by Engle and Granger (Engle and Granger, 1987), Engle and Yoo (Engle and Yoo, 1987) and Johansen (Johansen, 1988): among these, the Johansen cointegration test is the representative method that is widely accepted and used. In this study, we also used the Johansen cointegration test to test whether there were long-run stationary relationships between each of the variables, and when cointegration was observed, we applied the Johansen

Table 2
Results of unit root tests for sales data.

Test type	Augmented Dickey-Fuller (ADF)			Dickey-Fuller GLS (ERS)		
Test for unit root	Level	1st difference	2nd difference	level	1st difference	2nd difference
Series						
Netbook (worldwide)	−2.580	−4.149***	−6.319***	−0.823	−4.218***	−6.442***
Wii (worldwide)	−0.413	−12.670***	−5.167***	−1.883*	−11.872***	−0.460
Wii (U.S.)	−0.078	−11.398***	−5.948***	−0.211	−11.588***	−12.931***
Wii (Japan)	−1.523	−8.181***	−12.677***	−0.639	−1.994**	−1.611*

Performed using Eviews 7.0.
* $P < 0.1$.
** $P < 0.05$.
*** $P < 0.01$.

Table 3
Results of unit root tests for search traffic data.

Test type	Augmented Dickey-Fuller (ADF)			Dickey-Fuller GLS (ERS)		
Test for unit root	Level	1st difference	2nd difference	level	1st difference	2nd difference
Series						
Netbook (worldwide)	−4.032***	−2.022	−1.854	−1.437	−2.064**	−1.896*
Wii (worldwide)	−2.859*	−3.455**	−25.131***	−4.218***	−1.529	−23.141***
Wii (U.S.)	−2.625	−3.505**	−24.783***	−3.706***	−1.452	−0.309
Wii (Japan)	−0.283	−2.188	−3.280*	−1.215	−0.645	−0.195

Performed using Eviews 7.0.
* $P < 0.1$.
** $P < 0.05$.
*** $P < 0.01$.

cointegration test to verify the correlation between search traffic and sales.

In this study, in preparation for performing the Johansen cointegration test, we first set up a VAR model in the levels (not the differences) of the data and then determined the appropriate maximum lag length for the variables in the VAR. If we set up an overly extended lag (order) length, this would reduce the serial correlation of the error term but it would also result in somewhat lowering the efficiency. Taking account of this trade-off, in this test we set the maximum lag (order) length for the VAR model lower than 8. Table 4 shows the cointegration test results we obtained by applying this lag length.

As seen in the test results on the cointegration between sales and searches for each product, shown in Table 4, we found one or more cointegration equations, confirming that searches and sales have a long-run relationship. In the analytical model, we chose to test the correlation between search traffic and sales, if a cointegration equation was found to be present in the cointegration analysis as discussed above, then differencing the raw data would result in a problematic loss of significant data. For this reason, we performed the VECM GCBEW tests without differencing the raw data. In other words, we performed the Granger causality test to confirm the lead/lag structure in regards to the causality which we already verified in cointegration.

Table 5 presents the VECM GCBEW test results for the searches and sales of products that have cointegration. Our results showed that in the optimal VECM, Netbook, Wii (worldwide) and Wii (U.S) were the cases in which we can reject the null hypothesis of excluding search traffic variables from the sales products equation. In other words, in the case of Netbook, Wii (worldwide) and Wii (U.S), lagged search traffic can have a significant effect relevant to the explanation of sales. In the only case of Wii (Japan), we failed to reject the null hypothesis of excluding search traffic variables from the sales products equation at the significance level of 0.1.

Synthesizing these results, we found that search traffic and sales exhibited statistically significant association in the majority of our case studies, with the exception of Wii (Japan). We also confirmed the possibility that search traffic preceded sales: based on Granger causality (GCBEW tests), we confirmed the possibility that there can be a lag in sales of around four to eight quarters. By contrast, it was found that Wii searches in Japan do not have a significant effect on the sales of Wii in Japan. We found that Yahoo, rather than Google, is the dominant search site in Japan. Also, in the case of Japan, since Japanese is the language used by most consumers, we needed to analyze the query using a search term in Japanese. For this reason, in the following section, we compare search traffic for new IT technology by taking into account differences that exist across countries, languages, and search engines.

4.2. Comparison of the search traffic of different products

Fig. 4 presents the results of comparing the Google search traffic in the United States with the search traffic of Naver (which is one of the dominant and most frequently used web search engine in Korea) for the four products respectively. These graphs comparing the search traffic in the two countries for each product were similar in form in the majority of cases. As regards the peak of search traffic, the difference was unclear in the majority of products, but the search traffic for IPTV was found to show a delayed search cycle in Korea. Comparing the peaks of the search cycle for IPTV, we observed a delay of around 12 to 24 months. This shows that the timing may differ even in cases where there is a similar trend of changes in the interest in the same technology (product).

To verify the correlations between search traffic in the U.S. and Korea, we also performed the VECM GCBEW tests. As seen in the test results on the cointegration between searches in the U.S. and Korea, shown in Table 6, we found one or more cointegration equations, confirming that two countries' searches have a long-run relation, and therefore we did not show the unit root test results for the time series of searches. Taking account of this trade-off, in this test we set the maximum lag (order) length for the VAR model lower than 8. Table 6 shows the cointegration test results we obtained by applying this lag length. In addition, in the case of IPTV, because we found a delay of around 12 to 24 months in Fig. 4, we set the maximum lag length for the VAR model lower than 24.

Table 7 presents the VECM Granger causality (GCBEW) test results for the searches in the U.S. and Korea that have cointegration. Our

Table 4
Johansen system cointegration test results for Netbook and Wii.

Cointegration parameters	Unrestricted cointegration rank test (trace)					Maximum lag (order) length for VECM
	Hypothesized no. of CE(s)	Eigenvalue	Trace statistic	0.05 critical value	P-values	
Series						
Netbook (worldwide)	None[a]	0.541	21.237	18.398	0.020	5
	At most 1	0.067	1.745	3.841	0.187	
Wii (worldwide)	None[a]	0.806	39.675	18.398	0.000	7
	At most 1	0.083	2.003	3.841	0.157	
Wii (U.S.)	None[a]	0.880	55.192	18.398	0.000	7
	At most 1[a]	0.243	6.404	3.841	0.011	
Wii (Japan)	None[a]	0.710	29.665	15.495	0.000	7
	At most 1	0.049	1.157	3.841	0.282	

[a] Denotes rejection of the hypothesis at a significance level of 0.05.

Table 5
Results Granger Causality/Block Exogeneity tests for Netbook and Wii.

Dependent variable	Excluded variable	Chi-sq	df	Probability	Null hypothesis testing
Netbook (worldwide) sales	Netbook (worldwide) search	17.029	5	0.004	Reject
Wii (worldwide) sales	Wii (worldwide) search	162.587	7	0.000	Reject
Wii (U.S.) sales	Wii (U.S.) search	124.954	7	0.000	Reject
Wii (Japan) sales	Wii (Japan) search	9.119	7	0.242	Do not reject

results showed that in the optimal VECM, Blu-ray, Bluetooth and VoIP were the cases in which we can reject the null hypothesis of excluding search traffic variables from the sales products equation. In other words, in the case of Blu-ray, Bluetooth and VoIP, search traffic in the U.S. can have a significant effect relevant to the explanation of search traffic in Korea. In the only case of IPTV (lag length: 1), we failed to reject the null hypothesis of excluding search traffic in the U.S. from search traffic equation in Korea at the significance level of 0.1. By contrast, in the case of IPTV (lag length: 23), lagged search traffic in the U.S. can have a significant effect relevant to the explanation of search traffic in Korea at the significance level of 0.1.

Synthesizing these results, we find it notable that there were numerous cases in which the pattern was similar for the technology in question, in spite of differences in terms of country, language, and search engine. In Section 4.1, we were able to observe a correlation between the rate of market penetration (or sales) and the search traffic: the peak of search traffic was frequently observed to occur when the rate of penetration was around 40%. This means that the search traffic can serve as an indicator of when sales will slow down or approach a level of maturity. Utilizing these characteristics of the search traffic can assist us in analyzing the sales (or adoption of new technology. If the peak of search traffic is not yet observable, in other words, if the search traffic is continuing to increase, then we can explain that the market is still growing; on the other hand, if we can detect a decrease in the search traffic, we can anticipate that the market growth will slow down within a few years.

5. Discussion

As discussed above, analogy forecasting refers to the method of estimating future demands of a new product by comparison and analogy with the demand pattern of a similar product, its dissemination status, or cases from developed countries, when there is no available past market data for the product to be forecasted. Similarity and importance are critical when selecting a comparative dimension for the utilization of analogy forecasting (Martino, 1993). In regards to importance, this study has shown that search traffic performs well in explaining the adoption life cycle of a new technology product, and in regards to similarity, this study found that the patterns could be similar across different countries. In this section, We present and look into two cases of analogy forecasting mentioned in Section 2.1: one where sales data from the past are accumulated, and the other where there exist no historic data yet or the measurement of data is difficult to estimate. Above all, we provide a case to conduct the direct analogy of sales by regarding search traffic as a similar life cycle, and discuss the superiority and accuracy of search traffic-based analogy as supported by the case analysis. The second application case proposes the analogy method which compares search traffics in the market with no sales data by making use of

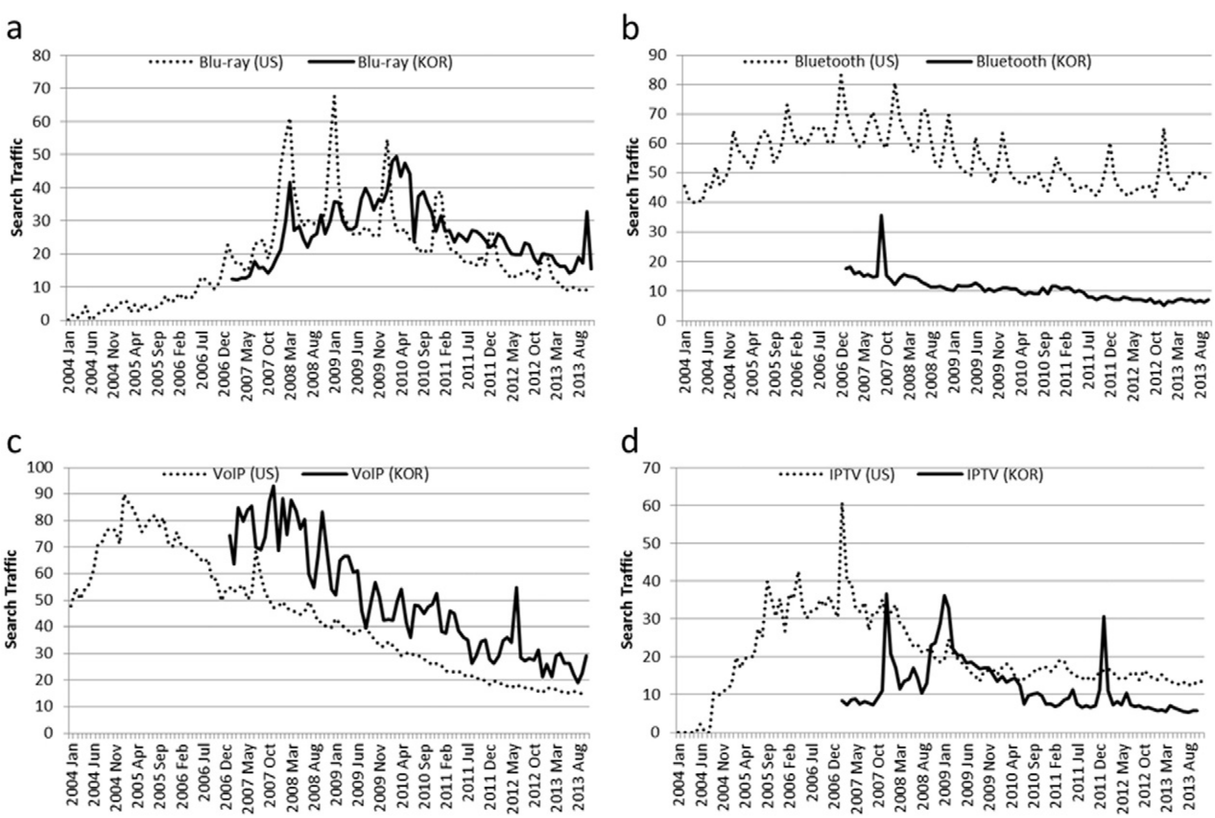


Fig. 4. Comparison of the monthly Google search traffic in the United States and Korea for Blu-ray (a), Bluetooth (b), VoIP (c) and IPTV (d).
Data source: Google (Google, 2016) and Naver (Naver, 2016).

Table 6
Johansen system cointegration test results for search traffic in the U.S. and Korea.

Cointegration parameters	Unrestricted cointegration rank test (trace)					Maximum lag (order) length for VECM
	Hypothesized no. of CE(s)	Eigenvalue	Trace statistic	0.05 critical value	P-values	
Series						
Blu-ray (U.S. & KOR)	None[a]	0.292	34.839	25.872	0.003	1
	At most 1	0.087	7.254	12.518	0.319	
Bluetooth (U.S. & KOR)	None[a]	0.396	59.799	25.872	0.000	1
	At most 1[a]	0.216	19.436	12.518	0.003	
VoIP (U.S. & KOR)	None[a]	0.310	30.650	15.495	0.000	2
	At most 1	0.017	1.393	3.841	0.238	
IPTV (U.S. & KOR)	None[a]	0.208	33.917	25.872	0.004	1
	At most 1[a]	0.178	15.533	12.518	0.015	
IPTV (U.S. & KOR)	None[a]	0.825	151.088	25.872	0.000	23
	At most 1[a]	0.597	51.797	12.518	0.000	

[a] Denotes rejection of the hypothesis at a significance level of 0.05.

the comparison between technology leaders and followers. Throughout the study of this case, we provide more in-depth discussion about the availability of search traffic in analogy forecasting.

5.1. Superiority and accuracy of the proposed model

The study discusses the excellence and effectiveness of search traffic-based analogy method. Prediction by a simple mathematical model would be more accurate and far easier if sales amount data of a product could be obtained by comparing the life cycle of the product mentioned in Section 4.1 with search traffic. In this aspect, in order to look into the competitiveness of search information in predicting future sales, we compared the accuracy of results from search traffic with that from a simple mathematical forecasting model regarding two products and four markets presented in Section 4.1, and selected the most common MSE (Mean Squared Error) indicator as a comparative one.

First, in case sales information in the market is already known, we applied the simplest analogy forecasting in order to see whether or not we could analogize a sales estimate from search traffic. Considering that search traffic is provided on a real-time basis, we applied the method of directly predicting the sales volume by the search traffic at the time of prediction. Just so as to consider the scale difference between two variables (i.e. search traffic and sales volume), we additionally considered only the proportion of the overall mean of the two variables (Analogy ideal model 1). We also compared together the results of analogical forecasting to the search traffic trends with the values over the preceding four quarters reflected, considering the delayed impact of search traffic indicated in Table 5 (Analogy ideal model 2). The six mathematical models compared with the two analogical forecasting methods mentioned above are as follows: Naïve seasonal, Simple average, Moving average, Exponential smoothing, Holt's model and Winters' model. The diffusion models such as regression models and Bass models were ruled out because it may result in bias effects about the selection of independent variables or similar products.

Let us briefly look into the six compared models. The basic Naïve model requires only historical values as the basis for forecasting. This method is best suitable for situations in which the data are stationary or in which any trends are regarded to be relatively stable. In this study, considering that the analyzed data are quarterly-based, we conducted the forecasting by utilizing the values over the preceding four quarters instead of the preceding quarter value. Simple average method is one that averages out historical values cumulatively to make predictions and is also advantageous in a stable time-series prediction. The study herein presented a cumulative mean value of sales volumes as the predicted value. Moving averages are most appropriate when the data are stationary and do not exhibit seasonality, where the time series analysis of quarterly data is conducted to calculate an estimate of the moving average value by the criteria of the previous four quarters. Next, exponential smoothing method is one of weighted moving average prediction techniques where the greatest weights are loaded onto the most recent data and the weights become decreasing geometrically as time passes. Historical data are essential to establish the best weighting factors in exponential smoothing. In this study, we used mathematical software to find the optimum weighting factors and applied the weighting factor to make the lowest mean-square-error (MSE) by simulations. Holt's model is also one of the exponential smoothing methods, and is appropriate where there exists a linearity trend but no seasonality. Our prediction was made by separately forecasting two kinds of weighting factors of the overall trend and the linearity trend. As similarly with exponential smoothing, we found two weighting factors that best reduce the MSE by simulations. Winters' exponential smoothing is utilized when the conditions for Holt's exponential smoothing come with seasonality, and thus predicts three weighting factors (Wilson and Keating, 2009). Along with the computer software, these three values minimize the MSE.

The results of comparing the accuracy of the eight prediction methods above are presented in Table 8. According to the results, search traffic-based analogy forecasting (analogy ideal model 1) is much more accurate than other sales volume-based forecasting methods for all of the four cases of the two products. Only the Winters' model with the compensation of the three weighting factors returned that its accuracy is higher than analogy forecasting. But the three ways including the Winters' model were recalculated by finding the weighting factors to minimize the MSEs by posteriorly simulating the prediction results to the overall life cycle. Meanwhile, we were able to confirm that analogy forecasting (Analogy ideal model 1 & 2) is competitive in terms of accuracy, considering that it only utilized the average of the two variables.

In Table 8, another analogy model is additionally adopted (Analogy real model) in consideration of potential practical utilization. The three kinds of exponential smoothing methods or the two analogy ideal models have the limitation that users must know weighting factors or scale differences exactly before the prediction procedure. That is, an unrealistic assumption is required that we should already know the distribution of the variables predicted prior to the forecasting. Hence, analogy method in realistic forecasting can consider only the search traffic at the moment of prediction and the mean values of both web search and sales volumes until the preceding quarter. The

Table 7
Results Granger Causality/Block Exogeneity tests for search traffic in the U.S. and Korea.

Dependent variable	Excluded variable	Chi-sq	df	Probability	Null hypothesis testing
Blu-ray search (US)	Blu-ray search (KOR)	4.459	1	0.035	Reject
Bluetooth (US)	Bluetooth (KOR)	5.772	1	0.016	Reject
VoIP (US)	VoIP (KOR)	18.001	2	0.000	Reject
IPTV (US)	IPTV (KOR)	2.148	1	0.143	Do not reject
IPTV (US)	IPTV (KOR)	32.432	23	0.092	Reject

Table 8
MSEs for the target products & forecasting models.

Products	Wii (U.S.)	Wii (JPN)	Wii (Worldwide)	Netbook	Remark
Forecasting models					
Analogy ideal 1	0.950	0.046	2.717	1.088	The mean value over the whole life cycle considered
Analogy ideal 2 - lag	0.995	0.154	4.110	9.327	The mean value over the whole life cycle and search 4Q lag considered
Naïve seasonal	1.141	0.077	3.446	9.568	
Simple average	2.599	0.173	9.445	11.358	
Moving average (4)	1.883	0.069	6.090	3.882	
Exponential smoothing	2.146	0.078	7.039	1.409	Simulated one weighted factor
Holt's model (exponential smoothing)	2.148	0.070	7.040	1.339	Simulated two weighted factor
Winters' model (exponential smoothing)	0.746	0.046	2.002	0.992	Simulated three weighted factor
Analogy real	1.032	0.066	2.856	3.339	Cumulative mean value up to the preceding quarter considered

prediction method of analogy real model can be expressed below, where s is the first time that sales are exposed, and w the first time of collecting web search, the values of both timings larger than the value $t-1$, respectively.

$$\text{Sales}_t = \text{Search}_T \times \frac{\sum_{i=s}^{t-1} \text{Sales}_i \div (t-1-s)}{\sum_{j=w}^{t-1} \text{Web Search}_j \div (t-1-w)} \tag{1}$$

The result predicted by the above formula is shown in the bottom of Table 8, where it maintains a high degree of accuracy to the three exponential smoothing methods although it is less accurate than Analogy ideal model 1 in general. We can regard that the reason why Analogy real model displayed high accuracy is that a similar effect like exponential smoothing appears while search traffic precedes sales volumes. It supports the above-mentioned claim that Wii surpasses Netbook in terms of accuracy of Analogy because Wii displays the feature of preceding search traffic relatively well.

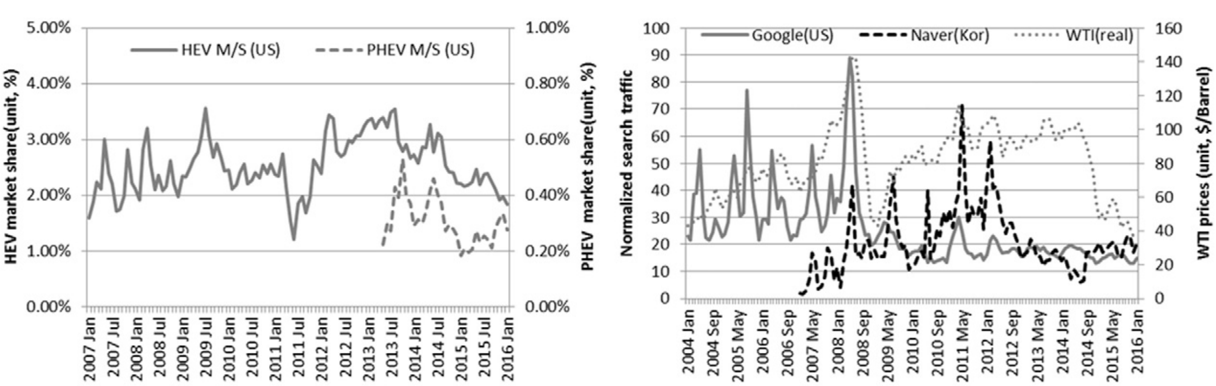
Taking consideration of the above results together, we identified that search traffic-based analogy forecasting proposed in the study is competitive in terms of both practicality and accuracy when compared with the usual mathematical prediction models.

5.2. The application example of web search based analogical forecasting

If there is real-time market data as in Section 5.1, a variety of mathematical forecasting models as well as analogy forecasting could be used and application of the analogy is also relatively easy. On the other hand, if sales data in the market is not given at all, application of mathematical forecasting models becomes more difficult and analogy forecasting gets much more competitive in the assessment (Goodwin et al., 2014). However, in order to conduct more accurate prediction, it is necessary to apply weighted analogy to various products (Goodwin et al., 2012) and structured analogy including consideration of multiple dimensions.

We present a simple case of analogy forecasting performed for the new PHEV market of a follower country, using the results above. The PHEV market in Korea is close to the state of a new market. Hyundai Motors Co. Ltd., a Korean domestic brand, launched the Sonata PHEV for the first time in late August 2015. This means that the PHEV remains unfamiliar in Korea and the market is close to a new state, and therefore there is hardly any market data. For this reason, we considered this to be a suitable case to which we can apply analogy forecasting. We selected HEV (Hybrid Electric Vehicle) as a similar product to analogize the demand pattern, and also selected the U.S. market, in which the sales of PHEVs began several years ago, as the advanced market to compare to the Korean market. The dimension with which to compare the two countries, Korea and the U. S., was the search traffic presented in this study. Multiple dimensions should be compared when performing structured analogies, but we regarded search traffic as a proxy variable which could directly explain the adoption of a new technology, comprehensively reflecting multiple dimensions such as revenues, oil prices, policies, etc. Just for the objective analysis of the search traffic as the main comparative dimension, we compare incomes in the economic dimension and patent applications in the technical dimension. The analogy methods in both leader and follower markets are summarized as follows.

1) Selection of prediction market (follower): PHEV market in Korea.
2) Selection of reference market (leader): PHEV market in the US.
3) Selection of analogy product: HEV.
4) Selection of main comparative dimension: Web search traffic.
5) Selection of sub comparative dimension: Income, Patent.
6) Comparison of dimension.
7) Forecasting of prediction market.



a) U.S. HEV & PHEV market share b) Search traffic and WTI price

Fig. 5. Comparison of the U. S. and Korean HEV markets for analogical forecasting (market share, search traffic and Oil (WTI) price).
Data source: HybridCars.Com (Wilson and Keating, 2009), EIA (EIA, 2016), Google (Jun and Park, 2016) and Naver (Naver_Trend, 2014).

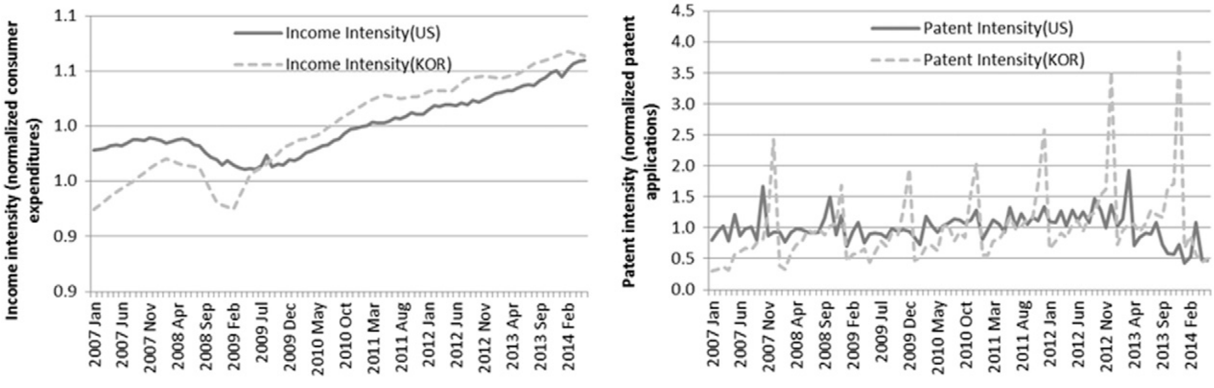


Fig. 6. Comparison of the U. S. and Korean HEV markets for analogical forecasting (income intensity, patent intensity).
Data source: BOK (BOK, 2016) and KIPRIS (KIPRIS, 2016).

According to the orders proposed above, we begin analogy forecasting. First, it is necessary to examine the data on the sales (or adoption) pattern of relevant products (or technologies) in developed countries. The sales market share (M/S) information for new HEVs and PHEVs since 2007 is shown on the left-hand side of Fig. 5. This M/S refers to the sales ratio of new cars that applied the technology in question, in comparison to the sales of new cars in general. The reason why the M/S ratio is used without directly providing the market size is because factors such as the difference in income between the leader and follower countries can result in differences in the size of vehicle demand in proportion to the population, etc. Therefore, we judged that it would be more appropriate to directly compare the M/S ratios, which take account of such factors.

Next, we compare search traffic to check similarities between the two markets. The U.S. HEV search traffic obtained from Google and the Korean search traffic from Naver are compared on the right-hand side of Fig. 5. The change in oil prices, which can have a strong influence on the sales of high fuel-efficient vehicles such as HEVs, is also presented on the right-hand side of Fig. 5 (Jun, 2012a; Jun, 2012b). We obtained the WTI (West Texas Intermediate) price by converting the nominal price into the real price, taking the consumer price index (CPI) into consideration. Based on to Fig. 5, we concluded that oil prices could be thought to have affected search traffic to some extent before 2008, but after 2008, the high oil prices did not influence high fuel-efficient vehicle search traffic in either country. In the second place, we compare subdimension while utilizing personal incomes as the economic dimension. Income intensity returns the compared value which is normalized by dividing the real consumer expenditures by the mean value in order to compare the trends between the two countries. In Fig. 6, there was no big difference in personal incomes in the United States and South Korea from 2007 to the mid-2014. Technological dimension has utilized the trend of patent applications related to fuel efficiency (patents that include "fuel efficiency" in their titles and abstracts). It also took advantage of the normalized values for the comparison between the two countries. In terms of the trends in patent intensity, there was a strong seasonality tendency especially in South Korea, but there was no big difference between the overall trends of both countries. Therefore, we assumed that search traffic trends as the main dimensions well describing the sales could have a dominant influence on the follower.

To forecast the Korean PHEV market based on the comparison above, we use HEV search traffic to compare the historical similarity of similar products or their life cycles in the two markets. The search traffic in Fig. 5 shows a difference of about 3 years measured from the peak. For statistical analysis, we performed unit root tests on the two sets of search traffic, but since the unit root was not found in the level variable, the search traffic of the two countries could not be observed to be non-stationary. Accordingly, we performed Pairwise Granger Causality tests without differentiation and found that the probability of causality for the two sets of search traffic was the highest in the case that lagged by 31 months (F-Statistic = 4.0788, Prob. = 0.0029). We thus concluded by analogy that the HEV markets in the two countries have a statistically significant similarity and that the U.S. market precedes the other by around 31 months. In conclusion, we analogized that the potential PHEV market in Korea would also have the time lag of around 31 months compared to the U.S. market.

In addition, we need to forecast the market size of the new technology or its M/S ratio. This is because the M/S ratios themselves can vary between the two countries due to differences in factors such as oil prices, consumer power, policy support (subsidy), etc. in both countries. There are several information sources for analogizing the market size, and although detailed data on the recent Korean HEV market is not available, it is known that the size of the Korean hybrid vehicle market is around 2% of the car market, from January to November in 2015 (Lee, 2016). When compared with the average (3.1%) of the U.S. HEV M/S ratio 31 months earlier, this confirms that in Korea, the M/S ratio of eco-friendly HEVs is only about 64% of the U.S. M/S ratio. Synthesizing the information above, we could predict the Korean PHEV market in the next 30 months as shown in Fig. 7. Considering that new vehicle market size in Korea is about 1.5 million or so annually, we can forecast the Korean PHEV market size in 2016 to be around 3500 vehicles or so.

Of course, the forecasting method and its results shown above do not guarantee continual forecasting. It would be stretched to forecast that the U.S and Korean PHEV markets will continue to show a difference of around 31 months and that the M/S ratio will stay around 64%. This is a problem that is common to all types of demand forecasting: in order to forecast continually using the same method, it is necessary to keep track of the relationship continually and analogize again. For instance, if the Korean PHEV market becomes more activated, we could replace the search word HEV with PHEV when comparing the search traffic across countries, and we may also conduct another comparison

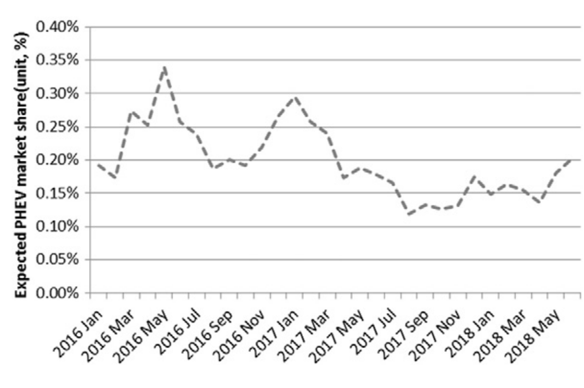


Fig. 7. The forecasts of Korean PHEV market by analogy.

of the differences in searches for HEV. In particular, although the U.S. market, for BEVs (Battery Electric Vehicles) has been already activated beyond that of PHEVs, BEVs have hardly been launched in Korea. This means that the potential demand related to BEVs shift to PHEVs, and leads to the growth of a market larger than the forecast given in Fig. 7. In addition, the life cycle gap with developed countries would be much further decreased by virtue of relevant policies or aggressive marketing in the country in which adoption occurred later, compared to the one in which adoption occurred earlier. In consideration of this aspect as well, it is expected that search traffic, which is relatively easy to measure, will serve as a useful variable.

While the United States is the first to benefit from the commercialized technologies as the leading market, the approaches above can also be enjoyed by other countries as well. If we compare with the countries which PHEV was disseminated at the early stage like Japan, search traffic is expected to be also a good dimension. However, it should be noted that Yahoo! Japan is the dominant search engine in Japan as mentioned in the results described in Table 5, and thus Google does not fit into the Japanese market. In addition, due to the difference of purchasing power, it is necessary to compare with different dimensions, as presented in this study. The present research methodology may also be applicable to market estimation of other follower countries than South Korea, but we need to consider language, search engine, etc. According to previous studies, if we consider the cases of Japan as well as the United States as the reference of leaders, more appropriate prediction is expected to be feasible. It can be explained in the same context with the reason why analogy method about plural products described before is necessary (Goodwin et al., 2012).

The reason why the present study strongly proposes search traffic to use analogy methods is fundamentally based on the advantage of easy examination of similarities. Thus, we need to give attention to the advantage that we enjoyed in conducting comparative analysis of search traffics between countries or between products much easier and quickly than that of other dimensions by utilizing search traffic. It is the ultimate goal of this study to contribute to both academic and practical realms in this regard.

6. Conclusion and limitation

Analogy forecasting is a qualitative method rather than a quantitative one. However, as explored in the above cases, if a variety of quantitative information such as search traffic based on Big Data, which has recently emerged as a focal point of attention, is used complementarily, analogy forecasting can be made more frequently applicable since it would become easier to identify the diverse correlations among multiple dimensions. Search traffic has already been used for various purposes in many fields, and this study specifically explored the potentials of web search traffic for helping us understand the adoption of innovative products by consumers. This study differs from preceding studies in that we analyzed whether search traffic appropriately explains the adoption (or sales) of a product in terms of the whole life cycle. While preceding studies limited their attempts to measure the search traffic to a single site, Google, and the consumers of one country, the United States, this study broadened the scope to another site, Naver, and the consumers of another country, Korea, and was thus able to verify that similar search traffic can be observed in other countries as well. Through this process, we have so far verified that search traffic is indeed a dimension that is well suited for comparison with analogical forecasting in terms of similarity and importance. Using search traffic, we were able to offer a case study of demand forecasting for PHEV, a new market in Korea.

This study underscores that search traffic has outstanding explanatory power for analyzing how consumers adopt new technologies or new products. Considering that the process of adopting new technology requires a relatively long time in comparison to conventional technological products and that information collection on such new technology may not translate as easily into their actual adoption, search traffic provides us with critically important information on how consumers think and behave during the adoption process (Tancer, 2008). The findings of this study therefore have the potential to contribute to greatly enhancing our objectivity and explanatory capability in regards to social phenomena by using search traffic. Furthermore, it can be used in technology forecasting and consumer behavior modeling for various fields such as marketing, thus contributing to the development of practical corporate strategies such as marketing tactics.

As emphasized by preceding studies, web search traffic has many advantages, allowing us to track consumer behavior in real time and to conduct research close to the population, and it can even offer forecasting capabilities. There are, however, also some clear disadvantages. A major problem with using search traffic is that trends can be driven by emotional factors such as excessive obsession with celebrities or fear, which can disseminate rapidly (Tancer, 2008). Preceding studies thus demonstrate that a temporary surge in interest can be manifested in search traffic even in response to negative news (Lui et al., 2011). This duality in the kinds of interest that generate searches is one of the reasons that we must exercise caution when interpreting search traffic.

Despite the benefits of using Google, which provides a large volume of information regarding raw data and research methodologies, this study ultimately relied on secondary data for the analysis of user search cycles and this can be regarded as one of its limitations. Another limitation is that Naver Trend, which was used for comparison with the data from Google, is still in the beta stage of service, and therefore its data or analysis results cannot yet be regarded as entirely reliable. Another problem is that although Google and Naver are currently dominant in the market, other sources of information such as SNS (Social Network Services) are also emerging, and there may also be a shift in the category of search engine users. Therefore, in the future it will be necessary to use a wider variety of sites to analyze the characteristics of the categories of users who are searching for information, and to include a larger number of cases when performing comparisons of search traffic by product as done in this study. In order to demonstrate that the implications and the conclusions drawn from this study are not based on "Data Snooping" and are validly generalizable, we will need more studies comparing various countries and products.

Acknowledgements

Seung-Pyo Jun, the first and corresponding author, acknowledges support from the Research Program (no. K16L04C02S04) at Korea Institute of Science and Technology Information. Furthermore, he thanks Sunhee Jung, Chanhyeok Jun and KISTI Technology Valuation Lab for the support of the research. Finally, these authors thank anonymous three reviewers for their constructive comments.

References

Bahmani-Oskooee, M., Alse, J., 1993. Export growth and economic growth: an application of cointegration and error-correction modeling. J. Dev. Areas 535–542.
Bayus, B.L., 1987. Forecasting sales of new contingent products: an application to the compact disc market. J. Prod. Innov. Manag. 4, 243–255.
Beatty, S.E., Smith, S.M., 1987. External search effort: an investigation across several product categories. J. Consum. Res. 14, 83–95.
Bettman, J.R., Park, C.W., 1980. Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: a protocol analysis. J. Consum. Res. 7, 234–248.
BOK, 2016. Korea and U.S. personal consumption expenditures. Bamk of Korea (http://ecos.bok.or.kr/).
Butler, D., 2013. When Google got flu wrong. Nature 494, 155.
Case, L., 2012. Are netbooks dead? The prognosis is grim. PCWorld.
Choi, H., Varian, H., 2009. Predicting initial claims for unemployment benefits. Google Inc.
Choo, C.W., Detlor, B., Turnbull, D., 1999. Information Seeking on the Web—An Integrated Model of Browsing and Searching.
comScore, 2013. August 2013 U.S. Search Engine Rankings. comScore (www.comscore.com).
Daim, T.U., Rueda, G., Martin, H., Gerdsri, P., 2006. Forecasting emerging technologies: use of bibliometrics and patent analysis. Technol. Forecast. Soc. Chang. 73, 981–1012.
Dignan, L., 2013. Netbook to die off 2015, says iSuppli. ZDNET.

EIA, 2016. Petroleum & other liquids. U.S. Energy Information Administration (http://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=rwtc&f=m).
Engel, J.F., Blackwell, R.D., Miniard, P.W., 1995. Consumer Behavior. 8 ed. The Dryden Press, New York.
Engle, R.F., Granger, C.W., 1987. Co-integration and error correction: representation, estimation, and testing. Econometrica 251–276.
Engle, R.F., Yoo, B.S., 1987. Forecasting and testing in co-integrated systems. J. Econ. 35, 143–159.
Fernández-Durán, J.J., 2014. Modeling seasonal effects in the bass forecasting diffusion model. Technol. Forecast. Soc. Chang. 88, 251–264.
Fox, B., 2002. Replacement for DVD unveiled. Newscientist. Com.
Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S., Brilliant, L., 2009. Detecting influenza epidemics using search engine query data. Nature 457, 1012–1014.
Goel, S., Hofman, J.M., Lahaie, S., Pennock, D.M., Watts, D.J., 2010. Predicting consumer behavior with Web search. Proc. Natl. Acad. Sci. 107, 17486–17490.
Goldenberg, J., Libai, B., Muller, E., 2002. Riding the saddle: how cross-market communications can create a major slump in sales. J. Mark. 66, 1–16.
Goodwin, P., Wright, G., 2010. The limits of forecasting methods in anticipating rare events. Technol. Forecast. Soc. Chang. 77, 355–368.
Goodwin, P., Dyussekeneva, K., Meeran, S., 2012. The use of analogies in forecasting the annual sales of new electronics products. IMA J. Manag. Math. 24, 407–422.
Goodwin, P., Meeran, S., Dyussekeneva, K., 2014. The challenges of pre-launch forecasting of adoption time series for new durable products. Int. J. Forecast. 30, 1082–1097.
Google, 2016. Google Trends.
Google_Trends, 2014. Google Trends Support.
Granger, C.W., 1988. Some recent development in a concept of causality. J. Econ. 39, 199–211.
Graves, L.E., Ridgers, N.D., Williams, K., Stratton, G., Atkinson, G.T., 2010. The physiological cost and enjoyment of Wii Fit in adolescents, young adults, and older adults. J. Phys. Act. Health 7, 393–401.
Green, K.C., Armstrong, J.S., 2007. Structured analogies for forecasting. Int. J. Forecast. 23, 365–376.
Grübler, A., Nakićenović, N., Victor, D.G., 1999. Dynamics of energy technologies and global change. Energy Policy 27, 247–280.
Hirunyawipada, T., Paswan, A.K., 2006. Consumer innovativeness and perceived risk: implications for high technology product adoption. J. Consum. Mark. 23, 182–198.
Hodgson, J., 2008. Cheap PCs weigh on Microsoft. The Wall Street Journal.
Jabłońska-Sabuka, M., Sitarz, R., Kraslawski, A., 2014. Forecasting research trends using population dynamics model with Burgers' type interaction. J. Informetr. 8, 111–122.
Järvenpää, H., Mäkinen, S., 2008. Empirically detecting the hype cycle with the life cycle indicators: an exploratory analysis of three technologies. Industrial Engineering and Engineering Management. IEEM 2008. IEEE International Conference on, IEEE 2008, pp. 12–16.
Johansen, S., 1988. Statistical analysis of cointegration vectors. J. Econ. Dyn. Control. 12, 231–254.
Jun, S.-P., 2012a. A comparative study of hype cycles among actors within the socio-technical system: with a focus on the case study of hybrid cars. Technol. Forecast. Soc. Chang. 79, 1413–1430.
Jun, S.-P., 2012b. An empirical study of users' hype cycle based on search traffic: the case study on hybrid cars. Scientometrics 91, 81–99.
Jun, S.-P., Park, D.-H., 2016. Consumer information search behavior and purchasing decisions: empirical evidence from Korea. Technol. Forecast. Soc. Chang. 107, 97–111.
Jun, S.-P., Yeom, J., Son, J.-K., 2014a. A study of the method using search traffic to analyze new technology adoption. Technol. Forecast. Soc. Chang. 81, 82–95.
Jun, S.-P., Park, D.-H., Yeom, J., 2014b. The possibility of using search traffic information to explore consumer product attitudes and forecast consumer preference. Technol. Forecast. Soc. Chang. 86, 237–253.
Kim, T., Lee, D.-J., Koo, S., 2014. Determining the scale of R&D investment for renewable energy in Korea using a comparative analogy approach. Renew. Sust. Energ. Rev. 37, 307–317.
KIPRIS, U.S. and Korean patent applications, 2016. Korean Intellectual Property Rights Information Service. http://www.kipris.or.kr/khome/main.jsp.
Koreanclick, 2012. The highest growth site (the first half of 2012 in Korea). Koreanclick (www.koreanclick.com).
Kotler, P., Keller, K., 2008. Marketing Management. 13th ed. Pearson Prentice Hall, Upper Saddle River, NJ.
Kumar, N., Lang, K.R., Peng, Q., 2005. Consumer search behavior in online shopping environments. System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on, IEEE, p. 175b.
Lee, S.-Y., 2016. Hybrid car, last year domestic sales up 34 991. Chosun Biz. ChosunBiz.com, Seoul Korea.
Lee, J., Lee, C.-Y., Lee, K.S., 2012. Forecasting demand for a newly introduced product using reservation price data and Bayesian updating. Technol. Forecast. Soc. Chang. 79, 1280–1291.
Lui, C., Metaxas, P.T., Mustafaraj, E., 2011. On the predictability of the US elections through search volume activity. Proceedings of the IADIS International Conference on e-Society, Citeseer.
Mackinnon, J.G., 1996. Numerical distribution functions for unit root and cointegration tests. J. Appl. Econ. 11, 601–618.
Mahajan, V., Muller, E., Bass, F.M., 1990. New product diffusion models in marketing: a review and directions for research. J. Mark. 1–26.
Martino, J.P., 1993. Technological Forecasting for Decision Making. McGraw-Hill, Inc.
Moe, W.W., 2003. Buying, searching, or browsing: differentiating between online shoppers using in-store navigational clickstream. J. Consum. Psychol. 13, 29–39.
Moon, B.-J., 2004. Consumer adoption of the internet as an information search and product purchase channel: some research hypotheses. Int. J. Internet Market. Advert. 1, 104–118.
Moore, G.A., 1999. Crossing the Chasm: Marketing and Selling High-tech Products to Mainstream Customers. Rev. ed. HarperBusiness, New York.
Naver, 2016. Nvaer Trend.
Naver_Trend, 2014. Naver Trend Help.
Nintendo, 2014. Consolidated Sales Transition by Region.
P.-L. To, Liao, C., Lin, T.-H., 2007. Shopping motivations on Internet: a study based on utilitarian and hedonic value. Technovation 27, 774–787.
Peres, R., Muller, E., Mahajan, V., 2010. Innovation diffusion and new product growth models: a critical review and research directions. Int. J. Res. Mark. 27, 91–106.
Pratt, K., 2010. Netbook, eReader, or iPad?–that is the question. Computers in New Zealand Schools. 22.
Raju, P.S., Lonial, S.C., Glynn Mangold, W., 1995. Differential effects of subjective knowledge, objective knowledge, and usage experience on decision making: an exploratory investigation. J. Consum. Psychol. 4, 153–180.
Rao, A.R., Sieben, W.A., 1992. The effect of prior knowledge on price acceptability and the type of information examined. J. Consum. Res. 19, 256–270.
Rogers, E.M., 2003. Diffusion of Innovations. Free Press, New York., NY.
Routley, M., Phaal, R., Probert, D., 2013. Exploring industry dynamics and interactions. Technol. Forecast. Soc. Chang. 80, 1147–1161.
Ruef, A., Markard, J., 2010. What happens after a hype? How changing expectations affected innovation activities in the case of stationary fuel cells. Tech. Anal. Strat. Manag. 22, 317–338.
SAS, 2012. New product analogy forecasting structured analogies. White Paper.
Shim, S., Eastlick, M.A., Lotz, S.L., Warrington, P., 2001. An online prepurchase intentions model: the role of intention to search. J. Retail. 77, 397–416.
Sproles, G.B., Geistfeld, L.V., Badenhop, S.B., 1980. Types and amounts of information used by efficient consumers. J. Consum. Aff. 14, 37–48.
Tancer, B., 2008. Click: What Millions of People Are Doing Online and Why It Matters. Hyperion.
Thelwall, M., 2008. Bibliometrics to webometrics. J. Inf. Sci. 34, 605–621.
Thelwall, M., Vaughan, L., Björneborn, L., 2005. Webometrics. Annual Review of Information Science & Technology 39 pp. 81–135.
Thomas, R.J., 1985. Estimating market growth for new products: an analogical diffusion model approach. J. Prod. Innov. Manag. 2, 45–55.
van Lente, H., Spitters, C., Peine, A., 2013. Comparing technological hype cycles: towards a theory. Technol. Forecast. Soc. Chang. 80, 1615–1628.
Vaughan, L., Yang, R., 2013. Web traffic and organization performance measures: relationships and data sources examined. J. Informetr. 7, 699–711.
Vosen, S., Schmidt, T., 2011. Forecasting private consumption: survey-based indicators vs. Google trends. J. Forecast. 30, 565–578.
Wei, J., Bu, B., Liang, L., 2012. Estimating the diffusion models of crisis information in micro blog. J. Informetr. 6, 600–610.
Wikipidia, 2014. The introduction of Blu-ray, Bluetooth, MP3, VoIP, LED TV, IPTV, Wii. Wikipidia (www.wikipidia.org).
Wilson, J., Keating, B., 2009. Business Forecasting with ForecastX. 6th ed. McGraw-Hill/Irwin, Boston.
Zimmer, J.C., Henry, R.M., Butler, B.S., 2007. Determinants of the use of relational and nonrelational information sources. J. Manag. Inf. Syst. 24, 297–331.

Seung-Pyo Jun is a principal researcher at the Korea Institute of Science and Technology Information (KISTI). Also, he is an associate professor of Science and Technology Policy in University of Science & Technology (UST). He received his Ph.D. from the Science and Technology Studies Program at Korea University. His area of interest includes the demand forecasting, emerging technology detecting and new technology adoption analysis. Recently he is focusing on the new technology adoption analysis using big-data.

Tae-Eung Sung is a senior researcher at the Korea Institute of Science and Technology Information (KISTI), also being affiliated with University of Science & Technology (UST) as an adjunct professor. He received his Ph.D. in electrical and computer engineering from Cornell University. His research interest lies in industry and market information analysis, technology valuation and commercialization, and technology innovation and management, stretching onto technology and product life cycles, and demand forecasting.

Hyun-woo Park is working for Korea Institute of Science and Technology Information (KISTI) as a technology economist (1991-Present) and for University of Science and Technology, Korea (UST, Korea) as a professor. He worked with San Francisco State University as a Visiting Fellow (1996–1997) and University of California, Santa Cruz as a Research Scholar (2008–2009). He received the B.S., M.S., and Ph.D. in International Business (1991) from Hong-Ik University, and Ph.D. in Science and Technology Studies (2007) from Korea University. In KISTI, he is working for the development of practical models and automated systems in the fields of technology innovation and management, technology valuation and commercialization, and scientometrics.