Technological Forecasting & Social Change 118 (2017) 270–280

Contents lists available at ScienceDirect

Technological Forecasting & Social Change

Science foresight using life-cycle analysis, text mining and clustering: A case study on natural ventilation

M. Rezaeian [a,*], H. Montazeri [b,c], R.C.G.M. Loonen [b]

[a] Faculty of Economics, Management & Accounting, Yazd University, Iran
[b] Building Physics and Services, Department of the Built Environment, Eindhoven University of Technology, The Netherlands
[c] Building Physics Section, Department of Civil Engineering, KU Leuven, Leuven, Belgium

ARTICLE INFO

Article history:
Received 18 April 2015
Received in revised form 9 February 2017
Accepted 21 February 2017
Available online 22 March 2017

Keywords:
Science foresight
Life-cycle analysis
Knowledge gap identification
Sensitivity analysis
Text mining
Wind catcher

ABSTRACT

Science foresight comprises a range of methods to analyze past, present and expected research trends, and uses this information to predict the future status of different fields of science and technology. With the ability to identify high-potential development directions, science foresight can be a useful tool to support the management and planning of future research activities. Science foresight analysts can choose from a rather large variety of approaches. There is, however, relatively little information about how the various approaches can be applied in an effective way. This paper describes a three-step methodological framework for science foresight on the basis of published research papers, consisting of (i) life-cycle analysis, (ii) text mining and (iii) knowledge gap identification by means of automated clustering. The three steps are connected using the research methodology of the research papers, as identified by text mining. The potential of combining these three steps in one framework is illustrated by analyzing scientific literature on wind catchers; a natural ventilation concept which has received considerable attention from academia, but with quite low application in practice. The knowledge gaps that are identified show that the automated foresight analysis is indeed able to find uncharted research areas. Results from a sensitivity analysis further show the importance of using full-texts for text mining instead of only title, keywords and abstract. The paper concludes with a reflection on the methodological framework, and gives directions for its intended use in future studies.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Effective management and planning of research and development activities require strategic allocation of available resources (Arroyabe et al., 2015; Berloznik and Van Langenhove, 1998). This issue manifests itself at different scales and plays a role in private companies and public authorities as well as academia. For example, individual scientists and research departments have a keen interest in spending their time and money in areas with potential for high impact (Kajikawa et al., 2008; Kostoff, 2008; Kostoff and Schaller, 2001; Leydesdorff et al., 1994; Ogawa and Kajikawa, 2015). Likewise, (inter)national governmental institutions seek to establish policy instruments (e.g. legislation and funding schemes) that give priority to development and application of innovative solutions with the highest positive contribution for society (Coccia, 2009; Kidwell, 2013).

Identification of such high-potential research and development areas is a challenging task. Making well-informed decisions requires detailed knowledge of past findings and current trends, and a deep understanding of emerging technology pathways (Leydesdorff et al., 1994; Yoon and Park, 2005). At the same time, it asks for a broad perspective to oversee future needs while identifying the opportunities that arise in neighboring research domains. The context in which such decisions are made is becoming increasingly complex because traditional science and engineering domains are getting more and more interconnected (Morillo et al., 2003; Porter and Rafols, 2009). In addition, the information that is documented in patents, reports and research papers continues to grow in size at an exponential rate (Kajikawa et al., 2008; Kostoff and Schaller, 2001; Larsen and Von Ins, 2010; Bengisu and Nekhili, 2006). The availability of input for research and technology planning can therefore be perceived as overwhelming, especially for decision-makers who are new to the field. The inability to properly analyze and comprehend all this information may lead to wrong recommendations and suboptimal priorities in research and development agendas.

Science foresight refers to the collection of analysis and prediction methods that can assist the development of a science vision in order to prepare for future challenges or needs in science (Martin, 1995; Martin, 2010). It has successfully been implemented in different fields, such as economy (Nassirtoussi et al., 2014), environmental science (Dubarić et al., 2011; Iniyan and Sumathy, 2003), foresight (Saritas and

* Corresponding author at: Department of Economics, Management and Accounting, Yazd University, P.O. box 89195 - 741, Yazd, Iran.
E-mail address: mina.rezaiyan@ymail.com (M. Rezaeian).

http://dx.doi.org/10.1016/j.techfore.2017.02.027
0040-1625/© 2017 Elsevier Inc. All rights reserved.

Burmaoglu, 2015; Su and Lee, 2010), health science (Pereira and Escuder, 1999; Abbott et al., 2014), politics (Coates, 1985), nano science and technology (Huang et al., 2011; de Miranda Santo et al., 2006; Robinson et al., 2007), and social science (Baloglu and Assante, 1999; Singh et al., 2007). The literature on science foresight covers a wide variety of qualitative and quantitative means for monitoring clues and indicators of evolving trends and developments (Coates, 1985). To facilitate successful science foresight analyses, it is clear that the methodology needs to be matched with e.g., the purpose of the study, the size and quality of the database, and the type of output that is expected. However, the available information about the relative effectiveness of different science foresight methods is very limited, and it is therefore difficult to support such decisions. In addition, most methods perform well at some, but typically not all aspects of science foresight. The potential of combining the positive sides of different science foresight methods into one overall framework has so far remained relatively unexplored.

The main objective of this paper is to develop and evaluate a three-step methodological framework that can be used to identify knowledge gaps and provide new insights into development directions of a well-defined technological field. Although we aim at wider applicability, in this paper, the framework is developed and demonstrated with respect to wind catchers; a sustainable natural ventilation system for buildings. This topic was specifically chosen because it is manageable in scope and size (i.e. the veracity of the results can be checked), yet has experienced a complex development history, is an active field with mixed research methods, and has a non-trivial future outlook. This paper uses a combination of existing methods: life-cycle analysis, text mining and cluster analysis, but combines them in a novel way that has not been described before. Given the importance of the impact of textual data on the accuracy of text mining, a sensitivity analysis is also carried out for three cases, when (i) title, (ii) title, abstract and keywords, and (iii) full-text of the papers are considered as the textual data. This evaluation is based on the methodology of the research papers.

This paper continues by describing the development of a methodological framework for science foresight on the basis of life-cycle analysis, text mining and cluster analysis (Section 2). The sensitivity analysis is also presented in this section. Characteristics of wind catchers, the topic of the application study, are introduced in Section 3. In Section 4, this methodology is applied in the case of wind catchers, to describe the status of research in this field, predict future trends and identify knowledge gaps in order to identify possible opportunities for new research and development activities. In Section 5, a reflection on the methodological framework and its potential in future studies is given.

2. Methodology

2.1. Life-cycle analysis

Life-cycle analysis is a widely-used data analysis technique that can be applied to describe the historical development of a technology or research domain, and, subsequently, to estimate the future trend or perspectives. Ernst (Ernst, 1997) suggests that the accumulation of patent applications is useful for measuring technology trends. The evolution over time can be plotted as S-shape curve to represent its technology life-cycle. There are four stages in a technology life cycle: introduction, growth, maturity and saturation (Ernst, 1997). During the introduction stage, there is a little growth in the number of patent applications. The growth stage, on the other hand, is characterized by exponential growth. As the patent application rate declines, the maturity stage is entered. The saturation stage indicates limited growth with only few additional patent applications (Trappey et al., 2011a).

If the current stage of a science or technology is known, it would be possible to forecast the future trends and predict the saturation level and therefore, estimate the potential of the field for further and deeper studies. Knowledge about the maturity and future growth potential of science or technology innovations helps researchers, for example, to decide whether to continue investing resources or switch research directions (Trappey et al., 2011a; Campani and Vaglio, 2014; Trappey et al., 2010).

In this study, cumulative paper publications are used for predicting future development trends using Loglet analysis. The analysis is performed using "Loglet Lab" software. It refers to the decomposition of growth and diffusion into S-shaped logistic components, roughly analogous to wavelet analysis, popular for signal processing and compression (Meyer et al., 1999).

Eq. (1) presents the equation to calculate the logistic growth:

$$N(t) = \frac{K}{1 + exp\left[-\frac{\ln(81)}{\Delta t}(t - t_m)\right]} \tag{1}$$

where K is the asymptotic limit that the growth curve approaches and shows the saturation level of the growth, Δt is the characteristic duration that specifies the time required for a trajectory to grow from 10% to 90% of the limit K and t_m is the midpoint of the growth trajectory (Fig. 1).

First, the logistic growth is visualized by simply plotting data on an absolute and linear scale. The Fisher-Pry transform is used to transform the logistic curve into a linear one. By doing so, Δt, t_m and K can be determined. Further information is presented by Meyer et al. (Meyer et al., 1999).

Many growth and diffusion processes consist of several sub processes. Systems with two growth phases are called "bi-logistic". In such models, growth is the sum of two discrete wavelets, each of which is a three-parameter logistic, as presented in Eq. (2).

$$N(t) = N_1(t) + N_2(t) \tag{2}$$

2.2. Text mining

One of the most popular methods for science foresight is text mining. Text mining is used to identify valuable information such as relations, patterns or trends in textual data (Choudhary et al., 2009; Delen and Crossland, 2008; Ghazinoory et al., 2013). For example, it has been widely adopted to explore the complex relationships among scientific documents (de Miranda Santo et al., 2006; Singh et al., 2007). A main theme supporting text mining is the transformation of text into numerical data. This transformation uses statistical methods to convert text mining into a classical data mining encoding. Despite the inability to explicitly understand linguistic concepts such as grammar or word meaning, statistical text mining has proven remarkably successful (Weiss et al., 2010). Many projects have used different techniques of statistical text mining in various fields of science or technology. In these studies, the full-text or abstract of papers or patents are considered as the database. Table 1 provides an overview of previous studies in which

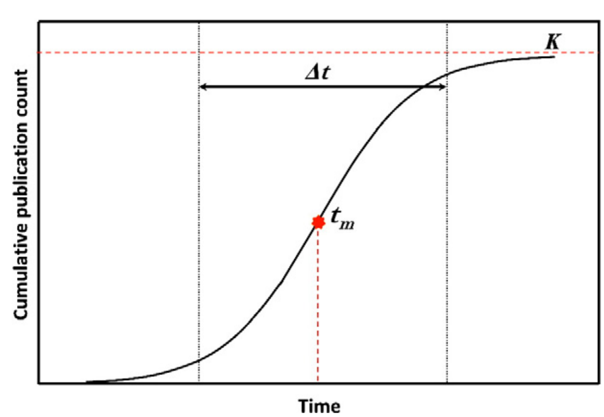

Fig. 1. A logistic curve and its parameters.

Download English Version:

https://daneshyari.com/en/article/5037051

Download Persian Version:

https://daneshyari.com/article/5037051

Daneshyari.com