Technology Analysis & Strategic Management

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

A visual scanning of potential disruptive signals for technology roadmapping: investigating keyword cluster, intensity, and relationship in futuristic data

Jieun Kim, Yongtae Park & Youngjo Lee

TECHNOLOGY ANALYSIS & STRATEGIC MANAGEMENT, 2016
http://dx.doi.org/10.1080/09537325.2016.1193593

A visual scanning of potential disruptive signals for technology roadmapping: investigating keyword cluster, intensity, and relationship in futuristic data

Jieun Kim[a], Yongtae Park[b] and Youngjo Lee[c]

[a]Data Science for Knowledge Creation Research Center, Seoul National University, Seoul, Korea; [b]Department of Industrial Engineering, School of Engineering, Seoul National University, Seoul, Republic of Korea; [c]Department of Statistics, Seoul National University, Seoul, Korea

ABSTRACT

Technology roadmapping of today's era is necessarily based on comprehensive scanning of various signals with the disruptive potential in future paths of market, product, and technology. Previous attempts of data-driven technology roadmaps have mainly focused on data from such sources as patents and literatures. However, as these sources catalogue posteriori trends of evolution, roadmaps based on these data cannot be counted on to predict disruptions. In this regard, futuristic data in technology foresight websites may provide a better source. The objective of this research, in response, is to develop a support system for technology roadmapping that uses futuristic data. To this end, we suggest keyword-based visual scanning approach involving three keyword maps, used in succession: keyword cluster map, keyword intensity map, and keyword relationship map. Particularly, keyword intensity map is designed using weak signal theory which can help identify the visibility, diffusion, and interpretation of signals.

ARTICLE HISTORY
Received 28 December 2015
Revised 19 April 2016
Accepted 17 May 2016

KEYWORDS
Technology roadmapping; futuristic data; weak signal; visual analysis

1. Introduction

Today's shortened technology lifecycles, hyper-competitions, and uncertain business environments present disruptive forms of innovation. Disruptive innovations that stimulate new forms of competitive dynamics and business models are distinct from sustaining types of innovation that do not upset existing industry patterns (Drew 2006). The prerequisites for building technology strategy for disruptive innovation have been discussed in literatures (Christensen et al. 2006). First, organisations need foresight into the possible future paths of technology innovation and their accompanying uncertainties. Second, organisations should have the capacity to absorb new knowledge and to make sense of disruptive signals of impending change from periphery (Hamel 2002; Christensen et al. 2006). This can be achieved by exploring multiple sources of expert advice and analysis of developments in business environments to develop creative thinking (Drew 2006).

To meet these prerequisites, technology roadmapping, the process for exploring and communicating the evolution of markets, products, and technologies, together with the linkages and discontinuities between various perspectives (Phaal, Farrukh, and Probert 2004), is increasingly considering various data sources. Previous attempts have proposed data-driven or keyword-based technology roadmaps (TRMs). They typically have depended on science and technology (S&T) data sources such as patents and literatures (Kostoff, Boylan, and Simons 2004; Lee et al. 2008; Martin and Daim

CONTACT Yongtae Park parkyt1@snu.ac.kr

2008; Geum et al. 2014; Jeong and Yoon 2015). S&T data can offer valuable information on the patterns of technological trajectory; however, they are fundamentally limited in responding to disruptive innovations. First, their scope is limited somewhat to 'posteriori trends', not future trends. This makes firms tend to support continuous or sustaining innovation only, which is based on existing technology evolution. Since the discontinuities and disruptions show a significant departure from the past or from the smooth extrapolation based on the past (Ansoff 1975; Kostoff, Boylan, and Simons 2004), the anticipation based on posteriori trends often fails, forcing firms to confront unfamiliar threatening or opportunistic events. Second, S&T data are limited to 'technology' viewpoints. Although the disruptive innovation stems from the emergence of a disruptive technology (Kostoff, Boylan, and Simons 2004; Walsh 2004), actual 'disruption' occurs when a disruptive technology is applied to the business model, that is, new product and market needs (Paap and Katz 2004; Christensen et al. 2006). Thus, signs of disruption should be scanned from integrated information of technologies, products, and markets.

In this regard, futuristic data from technology foresight websites may provide a remedial opportunity and be exploited as a potential source of technology roadmapping (Cachia, Compañó, and Da Costa 2007; Pang 2010; Schatzmann, Schäfer, and Eichelbaum 2013). Due to the advancement of the Internet, meaningful futuristic databases are accumulating in form of as technology foresight websites provided by technology-leading companies, professional foresight companies, national foresight organisations, futurologist communities, virtual communities of the general, social networking services, etc. These websites include future-oriented information such as opinions, forecasts, and trends. Thus, *futuristic data*, in this paper, is defined as a collection of future-oriented opinions extracted from online communities of large participation and collaboration of many experts and general populace (Kim and Park 2014). They contain information regarding a priori trends of technology innovation; they encompass easily extractable information on trends not only of technology but also of society, markets, and products that can be incorporated as components of technology planning; and they contain collective intelligence. According to Walsh (2004), a roadmap for a disruptive innovation must encompass a wide range of stakeholders and be more active and prescient than existing roadmapping efforts that focus on sustaining technology. Thus, futuristic data can be an effective source for TRMs.

The primary objective of this research, in response, is to develop a supporting system for technology roadmapping, using futuristic data to account for disruptive innovation. To this end, we suggest a keyword-based visual analysis approach to exploring key trends and integrating foresights of multiple sources from the vast amount of futuristic data. Using text mining (TM) and information visualisation algorithms, three types of keyword maps are constructed in consecutive order. First, a keyword cluster map is for scanning potential disruption areas that appear in futuristic data. Second, a keyword intensity map is for assessing the strength of keywords and selecting potential disruptive signals based on planning viewpoints. Third, a keyword relationship map is for gaining insight into the likely development paths of disruptive innovation and constructing keyword-based TRMs.

The rest of the paper is organised as follows. Section 2 describes the background regarding futuristic data, keyword-based visualisation, and weak signal. Section 3 explains our proposed research framework of keyword-based visual analysis. Section 4 illustrates the utility of suggested methods by the case of wearable computing. Finally, Section 5 presents the conclusions of our work.

2. Theoretical background

2.1. Futuristic data as a source of TRM

With the development of information and communication technology (ICT) since the mid-1990', various phenomena, such as production of a massive information, intelligent search methods, and communication via networks, created a new field for exploring future technologies that had largely been an exclusive territory of experts. Collective intelligence approach, that is, collecting, analysing, and utilising the knowledge from the public, is now often suggested as an alternative to traditional qualitative research to expand the scope of the study.

The role of ICT in technology foresight has been further elaborated by researchers. For example, Pang (2010) identified the way in which the web and ICT enhance expert-based foresight as (1) social scanning to discover the patterns extracted from open source data of futurists or foresight communities, (2) prediction markets to induce the experts' participation and synthesise their opinion, and (3) reviewing forecasts to support extensive evaluation on the results of applying previous foresight methods. Cachia, Compañó, and Da Costa (2007) argued that online communities can operate as a huge brain-storming tool that tests and improves future concepts, ideas, and scenarios. The potential of online communities for foresight is evident in their ability to serve as (1) a trigger of creativity through interaction and communication between foresight practitioners and interested individuals, (2) an index representing trends and changes of emotional and social activities, and (3) the big picture drawn from collective intelligence of various individuals discussing the long-term future objectives.

In this context, futuristic data have emerged from various technology foresight websites; the examples of futuristic data are databases and wikis, news and blogs, social rating systems, collaborative scenarios, prediction markets, etc. The providers of such futuristic data can be global technology-leading companies such as IBM and GE, professional technology forecasting or consulting companies such as Gartner and McKinsey, futurologist or expert communities such as Next Big Future, World Future Society, trend reporting websites such as Science Daily and Engadget, and social networks such as The Future of Facebook Project, etc. The field of technology on which they are focused is mainly ICTs, but other fields are included such as bio, nano, and energy, and social trends. Futuristic data include various levels of future-oriented information such as the current trend, short-term forecasts, or long-term forecasts. The forms of information are various as well, including news, report, magazine, web post (blog), forum (tread-reply), etc. For a detailed description and real-world examples, refer to the works of Schatzmann, Schäfer, and Eichelbaum (2013) and Kim and Park (2014).

Most of the futuristic data are in electronic form and disorganised text documents. Due to their quantity (as big data) and quality, knowledge is difficult to find with manual analysis. Cachia, Compañó, and Da Costa (2007) indicate such shortcomings of data from online communities: (1) the uncertainty to go beyond the central subjects because of unstructured and non-hierarchical structure of the information management system, (2) ethical concerns in terms of privacy and information authority, and (3) the problem of data format that is unstructured, non-codified, and difficult to extract. The aforementioned opportunities and challenges around futuristic data lead to the need for an analytical tool to uncover and digest knowledge from futuristic data to implement technology roadmapping.

2.2. Weak signal theory

In order to identify the potential disruptive signals, this study grounds on weak signal theories (Hiltunen 2008; Mendonça, Cardoso, and Caraça 2012; Yoon 2012). Weak signals are the early signs for future disruptions, discontinuities, trends, or other emerging big changes (Rossel 2009; Saritas and Smith 2011). According to Ansoff (1975), the coiner of the term a weak signal is defined as

> seemingly random or disconnected pieces of information that at first appear to be background noise but which can be recognized as part of a larger pattern when viewed through a different frame or by connecting it with other pieces of information.

In general, typical examples of weak signals are messages and signs associated with early developments in technologies, societal innovations, conflicts, origins of conflicts, demographic shifts, new rivals, new regulations, etc. (Saritas and Smith 2011).

Citing from Ansoff, Hiltunen (2008) proposed the three dimensions of future sign: signal, issue, and interpretation. Specifically, (1) signal is the number and/or visibility of signals, (2) issue means the number of events or the diffusion of phenomenon to a variety of other units, and (3) interpretation is the receiver's understanding of future sign's meaning; for example, an organisational point of view of this can be the importance of the sign for an organisation in the future. A sign of where all dimensions are small is a weak signal and it strengthens to a strong signal when there is a rise in at least one of the dimensions. Thus, we utilise these dimensions to assess keyword strength. Using signal and issue dimensions, Yoon (2012) suggested the keyword-based weak signal detection approach in web news. He quantitatively measured the 'degree of visibility' (i.e. signal) based on keyword frequency and 'degree of diffusion' (i.e. issue) based on document frequency, and constructed keyword portfolio maps to discriminate weak signals and strong signals. Since the strengthening of weak signals into strong signals requires a time period (Mendonça, Cardoso, and Caraça 2012), the strength of signals can represent whether the signals are related to short-term or long-term future. These criteria are utilised in developing a keyword intensity map.

3. Research framework

The objective of this paper is to propose the keyword-based visual analysis approach to support technology roadmapping from futuristic data. The potential disruptive signals are scanned at the keyword level in this paper. The research process consists of five steps, as shown in Figure 1.

First, futuristic data are collected from the technology foresight websites using web crawling techniques. Second, futuristic data, which consist of unstructured textual documents, are formatted into an analysable *keyword-document matrix* (KDM) using TM. Third, *future topics* are identified from the keyword cluster map. Since the extracted keywords are too massive and fragmented, a cluster of keywords that frequently co-occur in documents can represent a homogeneous topic and help to interpret the meaning in text easily. Thus, this paper classifies keywords into clusters and identifies their respective future-related topics. Fourth, the keywords for each topic are assessed in terms of *strength* using the keyword intensity map. Based on Hiltunen (2008) and Yoon's (2012) conceptualisation, this paper assesses the intensity of keywords using the dimensions of future signals, and then classifies the keywords into short-term, mid-term, and long-term keywords. Lastly, the keyword-based TRMs are developed using the keyword relationship map. In this stage, keywords are classified into TRM

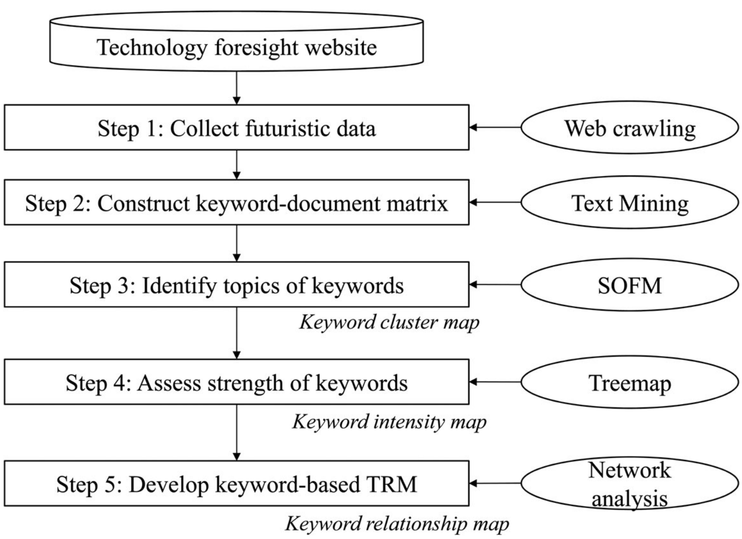

Figure 1. Research process.

metaphors that constitute the layers, that is, purpose, delivery, and resources; their *intra-layer and inter-layer relationships* are measured by similarity; and finally TRMs are constructed.

Steps 3–5 involve three types of keyword maps based on 'visual algorithms': keyword cluster map, keyword intensity map, and keyword relationship map. The *keyword cluster map* facilitates scanning future topics that appear and are discussed in futuristic data; in this purpose, we apply a self-organising feature map (SOFM) as a visual clustering algorithm. The *keyword intensity map* represents the keyword signals' visibility and diffusion in the purpose of classifying and selecting keywords as short-term, mid-term, and long-term keywords; for this, we utilise the *treemap* approach as a visual filtering method. The *keyword relationship map* uncovers the relationships among keywords using the *network analysis* approach. The map helps in gaining insight into the likely development paths of innovation. The detailed explanations of the proposed approach are provided in the following sections.

3.1. Collecting data

The data sources of this study are the futuristic data posted in technology foresight websites. Since boundaries of futuristic data are wide and a variety of information can be collected, technology strategists should carefully decide objectives (why we analyse futuristic data) and targets (what to focus in futuristic data). They must first decide the objective of planning (connecting with vision and mission) and the area of target technology (or market), and then identify and select relevant technology foresight websites. Futuristic documents would then be collected from the selected websites through the web crawling technique.

3.2. Constructing KDM

Futuristic data are collected as html documents consisting of unstructured texts. In order to build maps, the data should be processed into a structured format. TM starts with extracting terms from documents. Since there are meaningless keywords such as stop-words, the overall keyword set is refined by eliminating such keywords. Then, a KDM, as shown in Equation (1), is constructed. The row represents the keywords whereas the column denotes each futuristic document. The value of each cell in the matrix, kf_{ij}, denotes the normalised frequency of the ith keyword in the jth document ($i \in [1, n]$ and $j \in [1, d]$). Since the length of the document, that is, the sum of total keywords' frequency in the jth document, can vary, we utilise the keyword frequency divided by the length of documents as kf_{ij}. Row vectors of KDM are keyword vectors of d dimension.

$$\mathrm{KDM} = \begin{array}{c|c} & \begin{array}{ccc} \mathrm{period}_1 & \dots & \mathrm{period}_T \end{array}, \\ & \begin{array}{cccc} \mathrm{doc}_1 & \mathrm{doc}_2 & \dots & \mathrm{doc}_d \end{array}, \\ \begin{array}{c} \mathrm{keyword}_1 \\ \mathrm{keyword}_2 \\ \dots \\ \mathrm{keyword}_n \end{array} & \begin{bmatrix} kf_{11} & kf_{12} & \dots & kf_{1d} \\ kf_{21} & kf_{22} & \dots & kf_{2d} \\ \dots & \dots & \dots & \dots \\ kf_{n1} & kf_{n2} & \dots & kf_{nd} \end{bmatrix} \end{array}. \tag{1}$$

3.3. Identifying future topics: keyword cluster map

First, the keyword cluster map identifies the main topics discussed in the futuristic websites. Among many clustering algorithms, this study uses SOFM, a type of artificial neural network that is a general unsupervised tool for ordering high-dimensional data in a way that similar input patterns are grouped spatially close to one another (Kohonen 1998). Without knowledge in the inter-relationships,

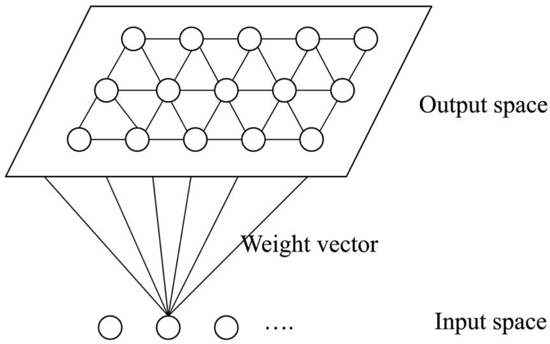

Figure 2. Structure of SOFM.

it produces a low-dimensional (typically two-dimensional), discretised representation of the space of the training data, or a map. As shown in Figure 2, an SOFM consists of neurons located on a two-dimensional grid arranged as various topologies such as rectangular or hexagonal lattices. Each neuron has two positions: one in input space (i.e. a weight vector) and another in output space on the map grid. Thus, SOFM is a vector projection method defining a nonlinear projection from the input space to a lower-dimensional output space. On the other hand, during the training, the weight vectors move so that they follow the probability density of the input data. Thus, SOFM is also a vector quantisation algorithm (Vesanto and Himberg 1999; Yoon, Yoon, and Park 2002). SOFM has been largely applied to information retrieval and knowledge discovery for textual documents (Morris et al. 2002) and to the technology management area such as technology planning using patents (Yoon, Yoon, and Park 2002; Park et al. 2013) and technology selection (Yu and Lee 2013).

In our case, using the keyword vectors as input data, keywords are mapped onto two-dimensional grids (i.e. by reducing dimensions from d to 2) as SOFM training, and then are clustered using k-means clustering. Since SOFM can conduct clustering and visualisation at the same time, or *clustering via visualisation* (Flexer 2001; Morris et al. 2002), the clustering process can be implemented in an interactive manner; for instance, after SOFM visualises the distribution of keywords in the 2D keyword space, the analyser can re-determine or refine both the shape and number of clusters until they fit the data density distribution and achieve homogeneity in meaning. With the keyword cluster map constructed as shown in Figure 3, topics and keyword sets assigned to the topics can be extracted.

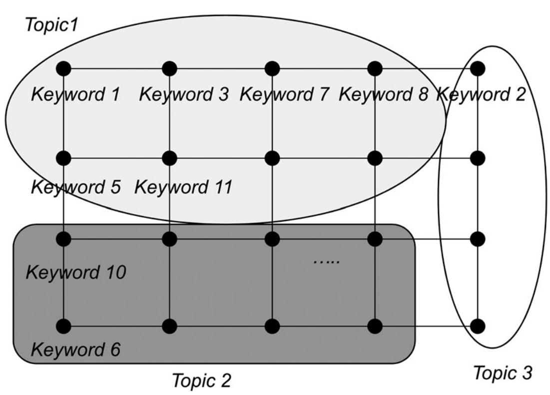

Figure 3. Keyword cluster map.

3.4. Assessing keyword strength: keyword intensity map

For each topic that pertains to the future, a keyword intensity map is constructed to assess keyword strength. As mentioned in Section 2.3, the map is based on the visibility, diffusion, and interpretation of weak signal theory. First, we measure the average keyword frequency (KF_i) (i.e. the average frequency of the ith keyword for all documents) and document frequency (DF_i) (i.e. the number of documents that include the ith keyword) and their average increasing rates between periods $t \in [1, T]$ as shown below in Equations (2) and (3):

$$\text{Average increasing rate of } \mathrm{KF}_i = \frac{1}{T-1}\sum_{t=2}^{T}\frac{\mathrm{kf}_i(t) - \mathrm{kf}_i(t-1)}{\mathrm{kf}_i(t) - \mathrm{kf}_i(t-1)}, \tag{2}$$

$$\text{Average increasing rate of } \mathrm{DF}_i = \frac{1}{T-1}\sum_{t=2}^{T}\frac{\mathrm{df}_i(t) - \mathrm{df}_i(t-1)}{\mathrm{df}_i(t) - \mathrm{df}_i(t-1)}, \tag{3}$$

where $\mathrm{kf}_i(t)$ is the average keyword frequency of the ith keyword in period t and $\mathrm{df}_i(t)$ is the document frequency of the ith keyword in period t. Using these variables two types of keyword intensity maps can be constructed as shown in Figure 4: keyword visibility map and keyword diffusion map. The maps can visualise the portfolio of keywords which are again classified into four respective quadrants. Keyword visibility map classifies keywords into four visibility categories: (1) developed sign that already gained visibility and therefore with low increasing rate, (2) developing sign that is actively obtaining visibility at a high increasing rate, (3) emerging sign with high increasing rate of visibility despite low visibility, and (4) undeveloped sign in which both rates are at a low. Keyword diffusion map, on the other hand, classifies keywords into four diffusion categories: (1) diffused, (2) diffusing, (3) emerging to diffuse, and (4) undiffused sign. For example, diffused sign means that the sign has already spread to many people and does not have a high, increasing rate of diffusion.

In Yoon (2012)'s framework, weak signals are identified at the intersection of the 'emerging to develop' and 'emerging to diffuse' quadrant whereas strong signals are at the intersection of 'developing' and 'diffusing'. However, there are other intersections such as 'emerging to develop' and 'undiffused', which also may potentially be weak signals or strong signals. Thus, to find the ambiguous distinction between weak signal and strong signal (Hiltunen 2008), this study reviews every combination of the quadrants in the visibility map and diffusion map, the blurry zone of weak signal and

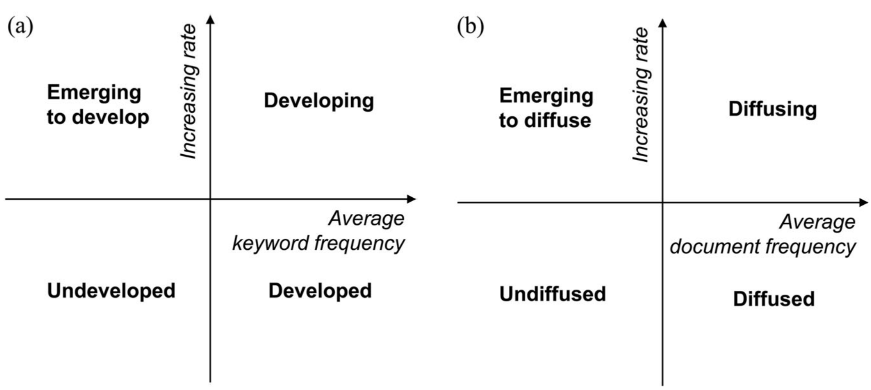

Figure 4. Conceptual definition of keyword intensity map. (a) Keyword visibility map. (b) Keyword diffusion map.

strong signal that potentially possesses disruptive signals. The combinations of developed-diffused and undeveloped-undiffused, which can be regarded as strong and weak signals with certainty, are excluded, as shown in the grey-coloured cells in Figure 5. In addition, since many studies argue that a weak signal strengthens into a strong signal through graduated stages requiring a time lag (Rossel 2009; Mendonça, Cardoso, and Caraça 2012), we assume that 'the category closest to weak signal (lower-right) will prevail in the long-term future, thus meriting consideration in long-term planning'. Based on this conceptualisation, the classification of keywords according to strength and time is shown in Figure 5.

Still, the classified keywords should be 'reassessed' in terms of organisational priorities because not all keywords would fit into an organisation's planning. This corresponds to the 'interpretation' dimension in Hiltunen's (2008) future sign model. However, interpreting those numerous keywords with various classes is too complex and time-consuming; as a solution, this paper suggests keyword intensity maps using a treemap algorithm, rather than a portfolio map (note that Figure 4 shows 'conceptual definition' of the map). Treemap is a space-filling approach for displaying hierarchical data by using nested rectangles (Shneiderman et al. 2012). Previous studies have suggested that hierarchical (tree) displays can potentially alleviate the problems of a traditional list display such as hypertexts, because they are more effective information access tools for browsing and information retrieval. In the technology management area, it has been applied to discovering innovation trajectories (Shneiderman et al. 2012), analysis of the patterns of technology adoption in management (Williams, Bernold, and Lu 2007), and project portfolio management (Cable et al. 2004).

So far, the analysis has derived the multi-variables for each keyword (i.e. topics, KF, increasing rate of KF, visibility category, DF, increasing rate of DF, diffusion category) in complex and hierarchical forms; a treemap can use visual coding such as hierarchical filling, size, or colour, to simultaneously display all such information. Figure 6 shows the suggested structure of a treemap-based keyword intensity map; in a keyword visibility map, for example, topic-visibility category-keyword hierarchy is represented by space-filling, the size of each keyword is set to KF, and the colour is coded as an increasing rate of KF. Although two maps are separated, they should be developed and read simultaneously for each topic in order to (1) categorise keywords into short-term, mid-term, and long-term,

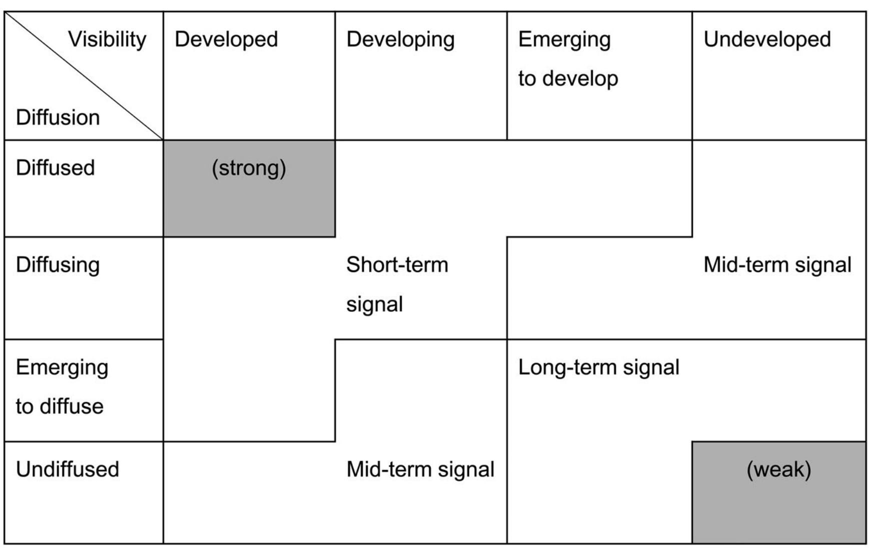

Figure 5. Classification of keywords according to keyword strength and time.

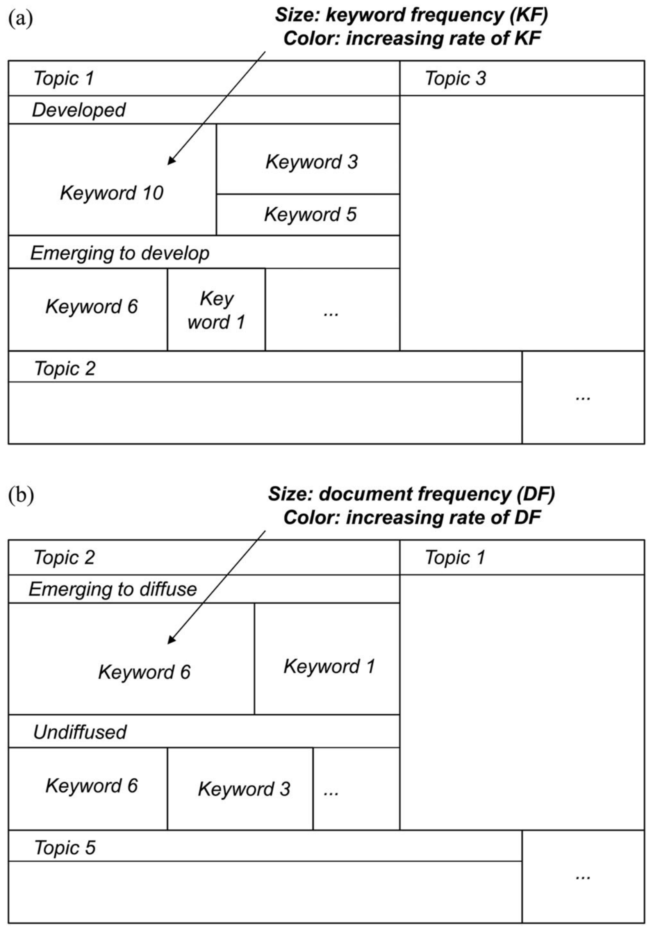

Figure 6. Keyword intensity map. (a) Keyword visibility map. (b) Keyword diffusion map.

as well as to (2) select strategically significant keywords. This process is supported by the 'drill-down interaction' of a treemap; if one clicks a topic, the keywords assigned within the topic are illustrated. Thus, the analyser can easily explore the keyword timelines by comparing two maps and select strategically important keywords in a comprehensive decision-making process.

Since Figure 5 uses the level of visibility and diffusion, the gap between visibility and diffusion can provide some general, but of course not absolute, guidelines for interpretation as shown below:

- Keywords of currently established trends, in the upper-left, are likely not disruptive in future because they are already pervasive; they should be filtered.
- Keywords in the lower-left section in Figure 5, with high visibility in spite of relatively low diffusion, are topics on which a small minority is intensely focusing their discourse; despite low diffusion, these topics can still be disruptive and should be carefully interpreted.
- Keywords in the upper-right section, with high diffusion and low visibility, may be everyday terms freely used by the majority. If this is the case, the keywords should be filtered.
- Garbage keywords, which should have been filtered in the preprocessing, may still be included; these keywords should be filtered.

3.5. Developing keyword-based TRMs: keyword relationship map

With the potentially disruptive keyword signals and their timelines identified, the next step is to uncover the relationships among keyword signals. To this end, the keyword relationship map is constructed using network analysis, a quantitative technique that facilitates the analysis of interactions, or 'edges' between actors, or 'nodes'. Network analysis has been widely used in S&T data mapping to analyse collaboration (Zheng et al. 2014), knowledge flow, or technology trends (Choi et al. 2013).

The keyword relationship map is constructed as follows. First, topics and keywords are reclassified as elements of strategic innovation planning. A typical example can be the three-fold TRM metaphors suggested by Phaal, Farrukh, and Probert (2004): purpose (know-why) (e.g. market, environment, business, trends, drivers, milestones, or strategy); delivery (know-what) (e.g. products, services, performance, features, systems, or risks); and resources (know-how) (e.g. technology, competence, partnerships, infrastructure, or R&D projects). Second, the relationships among keywords are measured using similarity measures. This study uses cosine similarity, a representative measurement for the similarity between two vectors of multi-dimension based on the cosine of the angle between them. The cosine similarity is represented using a dot product and magnitude, shown in Equation (4),

$$S_{ik} = \frac{\text{keyword}_i \cdot \text{keyword}_k}{|\text{keyword}_i||\text{keyword}_k|}, \tag{4}$$

where keyword_i and keyword_k are the row keyword vectors of the ith and kth keywords ($i, k \in [1, p]$) in KDM. Because the keyword frequency values are all positive, the cosine similarity of the two keywords ranges from 0 (independence) to 1 (exactly the same). Figure 7 shows the result of the measurement in matrix form. The sub-matrices of the same TRM metaphors describe intra-layer relationships; the rest denote inter-layer relationships.

With the information above, we can now construct the keyword relationship map. The Y-axis represents technology planning metaphors in TRM, and the X-axis represents keyword strength determined in Step 4, and keyword nodes are scattered across the map according to their classification and strength. Keyword edges that demonstrate high similarity scores are linked by lines. The model of this process is shown in Figure 8.

The keyword relationship map simplifies the identification of influence of one attribute on another, that is, identifying delivery options that fulfil a particular purpose of interest, providing knowledge on alternative resources that may be useful for a new delivery, or identifying new resources that can address particular delivery requirements.

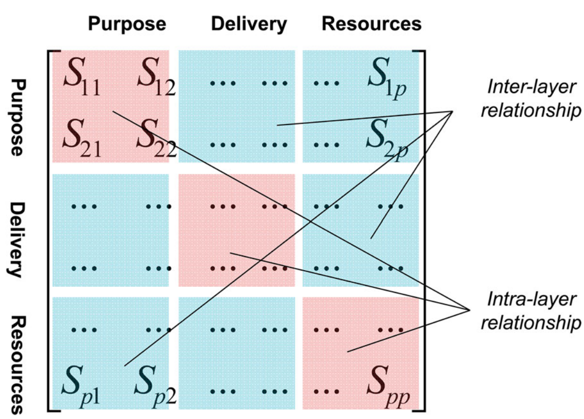

Figure 7. Similarity matrix.

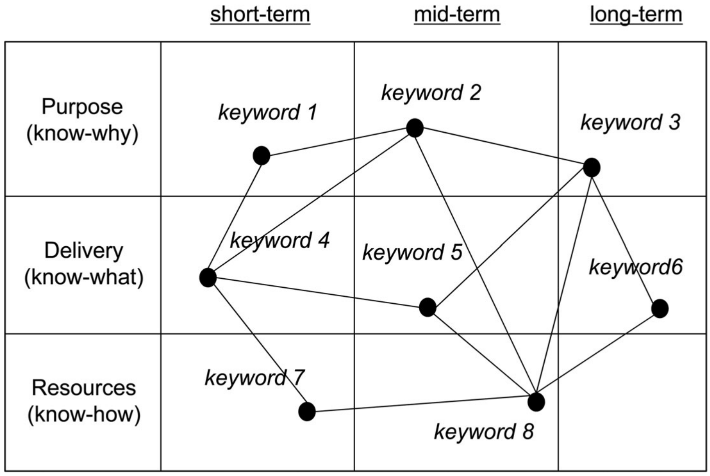

Figure 8. Keyword relationship map.

4. Illustrative case of wearable computing technology

The case study performs a keyword-based visual analysis of futuristic data on wearable computing technology, arguably the most prevailing disruptive technology of today. As convergence of information technology and biotechnology accelerates, wearable computing technology is predicted to revolutionise our everyday life; however, the uncertain shape of its future makes it difficult for strategic planners to identify disruptive signals that are both opportunities and threats in long-range planning. Visual scanning approaches can help organisations take advantage of these signals before their competitors by leveraging the predictions of various experts and the general population in technology foresight communities.

4.1. Data collection and preprocessing

First, we selected five technology foresight websites considering their relevance to the wearable computing field: Siemens (www.siemens.com/innovation/en/publications), MIT Technology Review (www.technologyreview.com/topics), Kurzweil Accelerating Intelligence (www.kurzweilai.net), World Future Society (www.wfs.org), and FutureTimeLine (futuretimeline.net).

Then, using two keywords, that is, ‘wearable’ and ‘wearable computing’, we searched for posts in the five websites in February 2014. Using the web crawling program by JAVA, we collected 454 futuristic documents as HTML files. To format the collected data into a structured KDM form, we applied TM by using TextAnalyst 2.1 software. As a result, 1777 keywords were extracted from 454 documents. After removing meaningless and low-frequency keywords (below five), 730 keywords were fixed. Then, the original keyword frequency was normalised according to the length of documents. Consequently, the KDM was constructed as a 730×453 matrix.

Note that elimination of low-frequency keywords is not a general guideline to be necessarily followed: although in this particular case study most keywords with frequency below five turned out to be garbage upon manual inspection, this is not always the case, and low-frequency keywords may in fact be helpful in expanding the source of a weak signal.

4.2. Identifying topics using an SOFM-based keyword cluster map

In order to identify topics, an SOFM-based keyword cluster map was constructed using ‘Matlab SOM Toolbox’ (Vesanto and Himberg 1999). By inputting the keyword vectors in KDM, SOFM was initialised and trained. We arranged the keyword neurons as a hexagonal lattice. From the iterative training of SOFM, the optimal size of grid in output space was determined as 13×10. Then, keywords were again clustered using a *k*-means clustering algorithm. In order to select the number of cluster (*k*), *k*-means was implemented with several values for *k* and their results were visually and quantitatively evaluated. For visual evaluation, we investigated the keywords of each cluster represented in SOFM, in terms of whether they are interpreted as a single, homogeneous topic. For quantitative evaluation, we utilised the Davies–Bouldin index (Davies and Bouldin 1979), which indicates the similarity of clusters with data densities, which is a decreasing function of distance from a vector characteristic of the cluster. Taken together, we identified 11 topics in Figure 9: 'medical and bionic', 'infotainment', 'workspace and office', 'nanotechnology', 'robotics for the handicapped', 'healthcare and neuroscience', 'industrial application', 'military robotics', 'electronics and interactive system', 'media', and 'fitness and wellness'. The data include particularly many keywords within the 'medical and bionic' topic, suggesting its importance in the future of wearable technology.

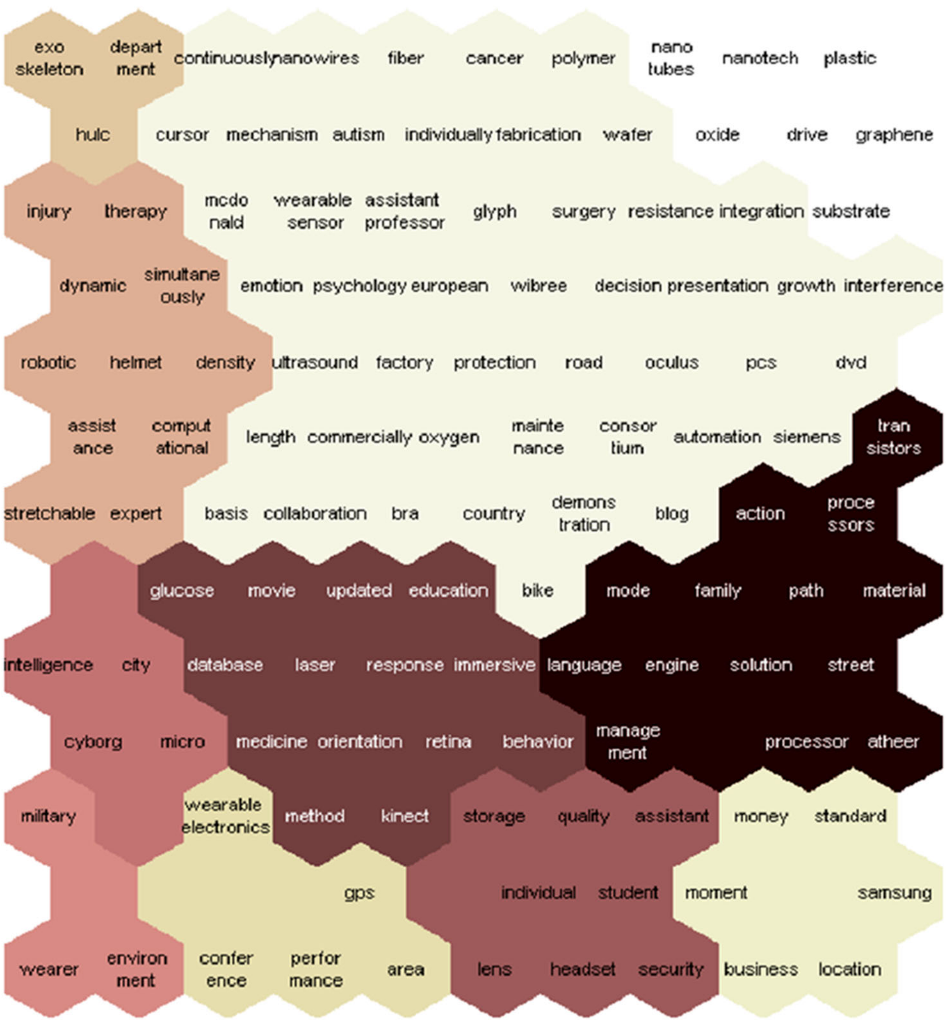

Figure 9. Result of building keyword cluster map.

4.3. Assessing the strength of keywords using a treemap-based keyword intensity map

To assess keyword strength, two types of keyword intensity maps were constructed as shown in Figure 10, using 'TIBCO®Spotfire' software. The keyword visibility/diffusion maps visualise all topics and their subsidiary keywords positioned according to their respective visibility/diffusion categories. Keyword/document frequency is represented by the varying size of the cells, and the average increasing rate of keyword/document frequency is represented by the varying shades of colour (red as min and green as max). As seen in the figure, keyword cells in 'undeveloped' and 'developed' categories are red due to their low increasing rate; 'developing' and 'emerging to develop' are green, with darker shades representing higher increasing rate. Compared to the portfolio-structure intensity maps in Appendix 1, the treemap structure of our maps allows simple and intuitive perception.

From Figure 10, we 'interpret' keywords that seem significant for wearable computing organisations. The maps allow for interactive interpretation by providing a drill-down function into lower levels. Figure 11 illustrates a representative example. Clicking on a topic (e.g. 'medical and bionic') returns all individual keyword cells within it. Clicking individual cells automatically highlights (as seen in grey) the equivalent cells of the other map (e.g. 'british' and 'ultrasound'). At this point, the analyser is ready for interpretation. The combination of 'developing' and 'undiffused' for the keyword 'ultrasound' qualifies it as a mid-term signal, and ultrasound's technical role to medical and bionic application of wearable computing is quite obvious; it is selected as a disruptive signal. Meanwhile, the keyword 'british' is categorised as 'undeveloped' and 'emerging to diffuse', and therefore a long-term signal; however, 'british' is a general term that is likely meaningless for medical and bionic applications, and is therefore filtered. Following this process, we extracted 241 disruptive signals from 730 keywords.

4.4. Developing TRMs using a network-based keyword relationship map

The selected disruptive keywords were re-classified into the TRM metaphor of strategic innovation planning. As an illustrative example, we selected 'workplace and office' to build the keyword relationship map as shown in Figure 12. We measured the cosine similarities among disruptive keywords, and utilised the networking software package 'UCINET 6' to visualise. The edges of keywords with similarity lower than the 0.05 cut-off were not linked. Here the X-axis is the timeline of the keywords coded by the shape of the nodes (i.e. up-triangle, square, and circle), whereas the Y-axis is the TRM layers coded by the colour of the nodes (i.e. yellow, orange, and green).

The results provide several interesting implications for TRM planning. The wearable computing field will seek rapidity (i.e. 'rapidly' in the purpose layer) in the workplace by developing 'analysis' and 'desktop' products and 'processor' technology in the short-term future. The development of 'processor' technology is associated with 'myo', an armband that provides wireless control with gestures and motion, in the mid-term, and 'myo' is related to the 'sunlight' purpose in the long term. Gesture control armbands like 'myo' are different from previous remote control processors in that they do not rely on cameras but on electrical activity in the users' muscles, and therefore are not subject to obstacles such as small workplace or the place with sunlight. The relationship suggests that the armband, presently confined to presentation or drone controlling, is expected to prevail in the mid-term and be applied to the sunlight environment in the long term. The keyword 'door' in the delivery layer is interesting: in the mid-term future, doors in workplaces will be changed into or associated with wearable products triggered by 'infrastructure' and 'server' technology, and will be applied for 'competition', 'success', 'connectivity', and 'official' purposes. Perhaps the relationship suggests that the office door can embed sensors to measure the status of passing people and be utilised as motivating bulletin boards or social networking boards.

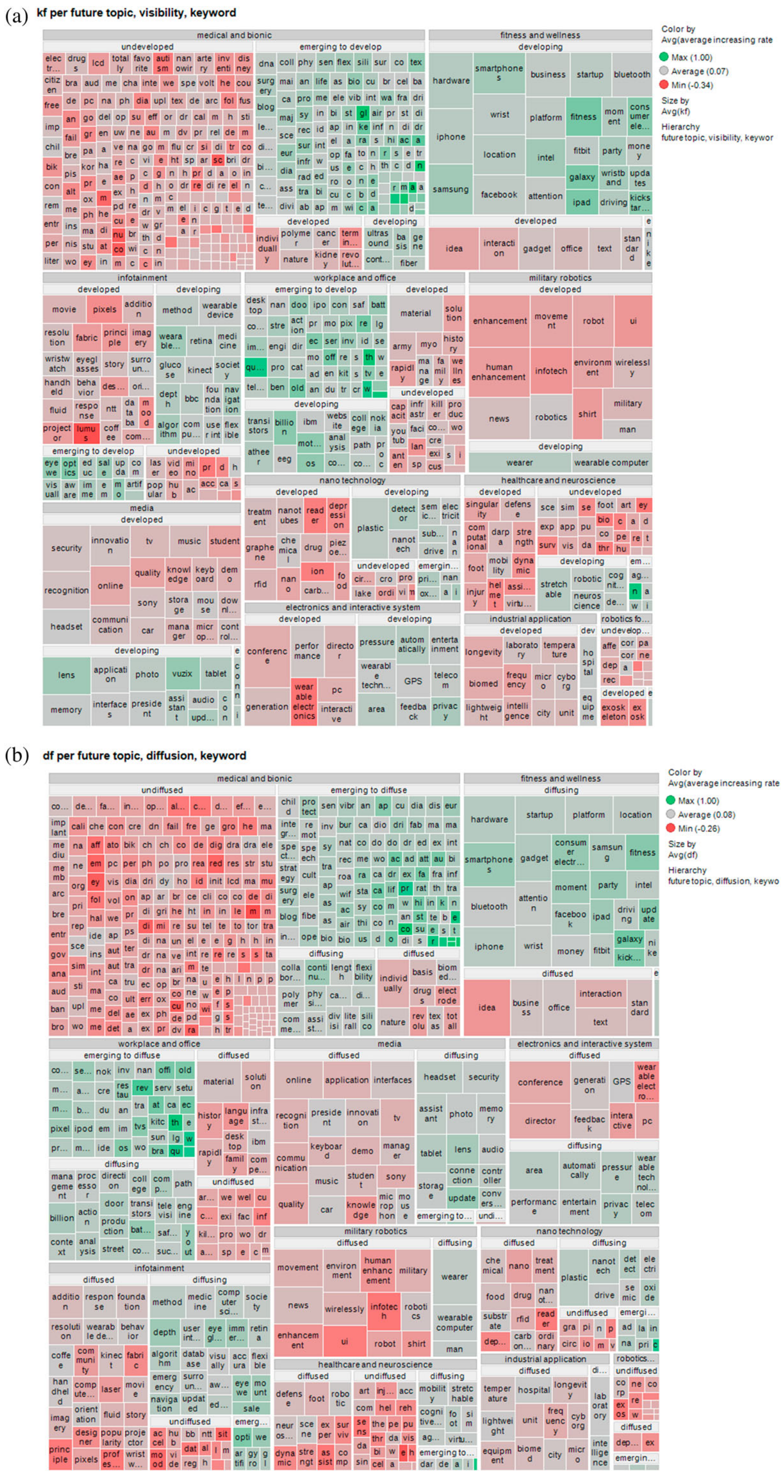

Figure 10. Result of building keyword intensity maps. (a) Keyword visibility map. (b) Keyword diffusion map.

Figure 11. Interactive drill-down example.

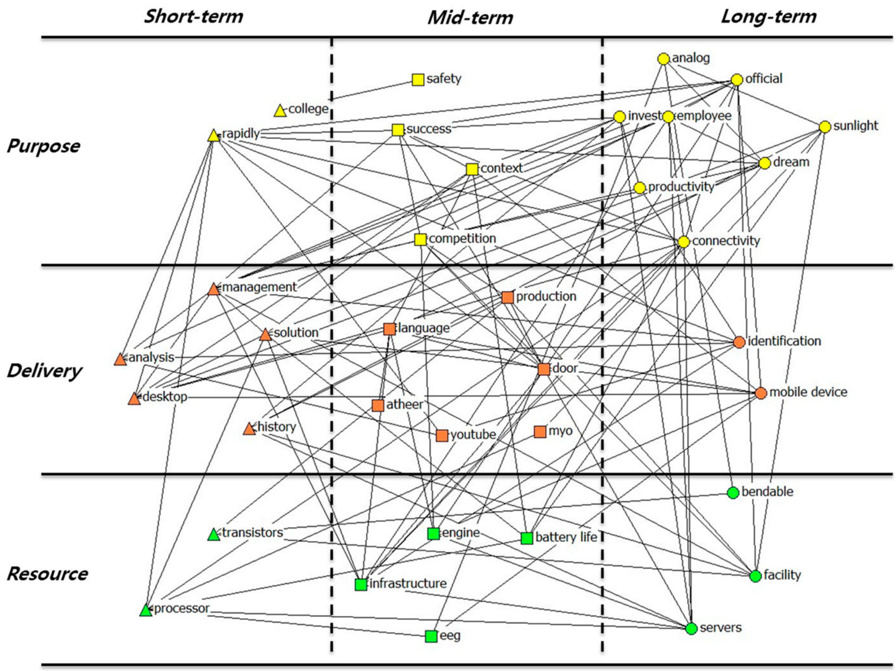

Figure 12. Keyword relationship map for 'workplace and office' topic.

5. Discussions and conclusions

5.1. Theoretical implication

This paper suggested a visual scanning approach to exploring potential disruptive signals from futuristic data and leveraging them into TRM. From a theoretical perspective, the proposed attempt can contribute to identifying opportunities and building foresight for disruptive innovation. The use of futuristic data can identify potential disruptive areas in current and future markets; examine facts, stimuli, and inspiration from many stakeholders' insights; discover disruptive applications of existing technologies and explore un-served market segments.

From a methodological perspective, the visual analysis approach can provide intelligence to roadmapping by processing vast amounts of futuristic data. First, using TM, the textual futuristic data are structured into analysable KDM. Second, the SOFM-based keyword cluster map identifies topics within diverse information from futuristic data as potential disruption areas in the future of technology. Because there is no information loss, the SOFM can retain all keywords and their divergent properties, including timelines, the drivers of market/business, key products, and technologies in the future trend. Third, the keyword intensity and relationship maps focus on the classification and selection of keywords. The treemap-based keyword intensity map explores the strength (visibility and diffusion), and the strategic importance (interpretation) of keywords in order to construct the timelines of TRM. Although the number of keywords can hinder clear cognition, the treemap's space-filling and drilling-down capability helps manage the complexity in analysis. The network-based keyword relationship map investigates the fields of keywords in order to construct the layers of TRM and intra-/inter-layer relationships. This offers basic evidence and support to technology planning.

The most important advantage of the framework is the improved efficiency and effectiveness for fast-developing TRM. The previous data-driven and keyword-based TRMs have focused on efficiency in terms of time and cost, and/or on objectivity compared to expert-based qualitative approaches. While we acknowledge the necessity of both efficiency and objectivity, however, we believe that in data-driven

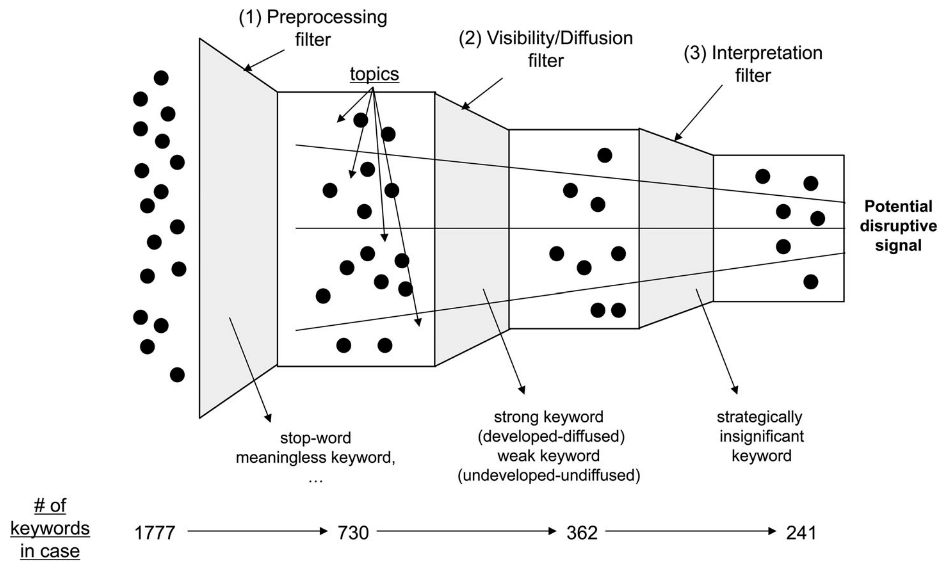

Figure 13. Keyword filtering funnel.

TRMs, expert (or manager's) knowledge often still plays a decisive role. Although keyword-based TRMs are apparently more efficient than the qualitative approach, they still require keyword selection and filtering processes that delay the TRM construction. Our 'interactive' visual analysis involves not only the data-driven information but also the analyser-driven knowledge: the analyser, with the help of the map, determines which keyword is to be focused on. The keyword selection process can be visually constructed as a filtering funnel, as shown in Figure 13: the preprocessing filter removes stop-words or meaningless keywords (step 2), the visibility/diffusion filter excludes overly strong or weak keywords (step 4), and the interpretation filter discards insignificant keywords (step 4). Before the second and third filters, keyword topics are divided so as to help the analyser perceive and select keywords. Taken together, these steps integrate data efficiency with human effectiveness.

5.2. Limitation and future studies

In spite of the contribution, there are rooms to be elaborated in future research. First, the research of futuristic data is in the infant stage, so the integration with S&T data can be further elaborated. Second, this study only considered the knowledge from outside the organisation; the factors from inside the organisation (e.g. vision, strength, weakness) can be integrated by using interactive visualisation functions. Lastly, the linkage between disruptive innovation and weak signal should be assessed further. Since there are arguments on the definition of and criteria for identifying a weak signal, our criteria based on keyword frequency, document frequency, and their increasing rates can be investigated. For instance, the factors for evaluating technologies (e.g. potential of acquisition, implementation, use, diffusion, securing, future, and technological impact) can be introduced to our treemap analysis. Finally, opinions vary on the assumption that a weak signal is realised in the long-term future; this assumption calls for further experimentation.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MOE and MSIP) [grant number 2014R1A1A2054064 and 2011-0030814].

Notes on contributors

Jieun Kim is a postdoctoral fellow in Data Science for Knowledge Creation Research Center at Seoul National University (SNU). She holds a BS and Ph.D. in Industrial Engineering from SNU. Her research interests are in technology planning and roadmapping, technology intelligence and visualisation.

Yongtae Park is a professor in the Department of Industrial Engineering at SNU. He holds a BS from SNU, and an MS and Ph.D. from the University of Wisconsin-Madison. His research topics cover various areas including technology innovation management and knowledge management.

Younjo Lee is a professor in the Department of Statistics at SNU. He has been recognised for his discovery of Hierarchical Generalised Linear Models and his research topics cover various areas including statistics, data mining, networks and communications.

References

Ansoff, H. Igor. 1975. “Managing Strategic Surprise by Response to Weak Signals.” *California Management Review* XVIII (2): 21–34. doi:citeulike-article-id:1109593.

Cable, John H., Javier F. Ordonez, Gouthami Chintalapani, and Catherine Plaisant. 2004. “Project Portfolio Earned Value Management Using Treemaps.” Proceedings of the Project Management Institute Research Conference, London.

Cachia, R., R. Compañó, and O. Da Costa. 2007. “Grasping the Potential of Online Social Networks for Foresight.” *Technological Forecasting and Social Change* 74 (8): 1179–1203.

Choi, Sungchul, Hongbin Kim, Janghyeok Yoon, Kwangsoo Kim, and Jae Yeol Lee. 2013. “An SAO-Based Text-Mining Approach for Technology Roadmapping Using Patent Information.” *R and D Management* 43 (1): 52–74. doi:10.1111/j.1467-9310.2012.00702.x.

Christensen, C. M., H. Baumann, R. Ruggles, and T. M. Sadtler. 2006. “Disruptive Innovation for Social Change.” *Harvard Business Review* 84 (12): 94–101.

Davies, David L., and Donald W. Bouldin. 1979. “A Cluster Separation Measure.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-1 (2): 224–227. doi:10.1109/TPAMI.1979.4766909.

Drew, S. A. W. 2006. “Building Technology Foresight: Using Scenarios to Embrace Innovation.” *European Journal of Innovation Management* 9 (3): 241–257.

Flexer, Arthur. 2001. “On the Use of Self-Organizing Maps for Clustering and Visualization.” *Intelligent Data Analysis* 5 (5): 373–384. IOS Press.

Geum, Youngjung, Hyeon Jeong Lee, Youngjo Lee, and Yongtae Park. 2014. “Development of Data-Driven Technology Roadmap Considering Dependency: An ARM-Based Technology Roadmapping.” *Technological Forecasting and Social Change* 91: 264–279. Elsevier Inc. doi:10.1016/j.techfore.2014.03.003.

Hamel, G. 2002. *Leading the Revolution: How to Thrive in Turbulent Times by Making Innovation a Way of Life*. Boston, MA: Harvard Business School Press.

Hiltunen, Elina. 2008. “The Future Sign and Its Three Dimensions.” *Futures* 40 (3): 247–260. doi:10.1016/j.futures.2007.08.021.

Jeong, Yujin, and Byungun Yoon. 2015. “Development of Patent Roadmap Based on Technology Roadmap by Analyzing Patterns of Patent Development.” *Technovation* 39–40 (May): 37–52. doi:10.1016/j.technovation.2014.03.001.

Kim, Jieun, and Yongtae Park. 2014. “Supporting Technology Foresight for Disruptive Innovation : Keyword-Based Visual Analysis for Futuristic Data.” Proceedings of the Asia Pacific Industrial Engineering & Management Systems Conference 2014, Jeju, South Korea.

Kohonen, Teuvo. 1998. “The Self-Organizing Map.” *Neurocomputing* 21 (1–3): 1–6. doi:10.1016/S0925-2312(98)00030-7.

Kostoff, Ronald N., Robert Boylan, and Gene R. Simons. 2004. “Disruptive Technology Roadmaps.” *Technological Forecasting and Social Change* 71 (1–2): 141–59. doi:10.1016/S0040-1625(03)00048-9.

Lee, Sungjoo, Seonghoon Lee, Hyeonju Seol, and Yongtae Park. 2008. “Using Patent Information for Designing New Product and Technology: Keyword Based Technology Roadmapping.” *R&D Management* 38 (2): 169–188. doi:10.1111/j.1467-9310.2008.00509.x.

Martin, Hilary, and Tugrul Daim. 2008. “Technology Roadmapping through Intelligence Analysis: Case of Nanotechnology.” *International Journal of Society Systems Science* 1 (1): 49–66.

Mendonça, Sandro, Gustavo Cardoso, and João Caraça. 2012. “The Strategic Strength of Weak Signal Analysis.” *Futures* 44 (3): 218–228. doi:10.1016/j.futures.2011.10.004.

Morris, Steven, Camille DeYong, Zheng Wu, Sinan Salman, and Dagmawi Yemenu. 2002. “DIVA: A Visualization System for Exploring Document Databases for Technology Forecasting.” *Computers & Industrial Engineering* 43 (4): 841–862. doi:10.1016/S0360-8352(02)00143-2.

Paap, Jay, and Ralph Katz. 2004. “Anticipating Disruptive Innovation.” *IEEE Engineering Management Review* 32 (4): 74–85. doi:10.1109/EMR.2004.25138.

Pang, Alex Soojung Kim. 2010. “Futures 2.0: Rethinking the Discipline.” *Foresight* 12 (1): 5–20. doi:10.1108/14636681011020191.

Park, Hyunseok, Kwangsoo Kim, Sungchul Choi, and Janghyeok Yoon. 2013. "A Patent Intelligence System for Strategic Technology Planning." *Expert Systems with Applications* 40 (7): 2373–23790. doi:10.1016/j.eswa.2012.10.073.

Phaal, Robert, Clare J. P. Farrukh, and David R. Probert. 2004. "Technology Roadmapping – A Planning Framework for Evolution and Revolution." *Technological Forecasting and Social Change* 71 (1–2): 5–26. doi:10.1016/S0040-1625(03)00072-6.

Rossel, Pierre. 2009. "Weak Signals as a Flexible Framing Space for Enhanced Management and Decision-Making." *Technology Analysis & Strategic Management* 21 (3): 307–20. doi:10.1080/09537320902750616.

Saritas, Ozcan, and Jack E. Smith. 2011. "The Big Picture – Trends, Drivers, Wild Cards, Discontinuities and Weak Signals." *Futures* 43 (3): 292–312. doi:10.1016/j.futures.2010.11.007.

Schatzmann, Jörg, René Schäfer, and Frederik Eichelbaum. 2013. "Foresight 2.0 – Definition, Overview & Evaluation." *European Journal of Futures Research* 1 (1): 1–15. doi:10.1007/s40309-013-0015-4.

Shneiderman, B., C. Dunne, P. Sharma, and P. Wang. 2012. "Innovation Trajectories for Information Visualizations: Comparing Treemaps, Cone Trees, and Hyperbolic Trees." *Information Visualization* 11 (2): 87–105. doi:10.1177/1473871611424815.

Vesanto, J., and J. Himberg. 1999. "Self-Organizing Map in Matlab: The SOM Toolbox." *The Matlab DSP Conference 1999*, 35–40. Espoo.

Walsh, S. T. 2004. "Roadmapping a Disruptive Technology: A Case Study: The Emerging Microsystems and Top-down Nanosystems Industry." *Technological Forecasting and Social Change* 71 (1–2): 161–185.

Williams, Trefor, Leonhard Bernold, and Huanqing Lu. 2007. "Adoption Patterns of Advanced Information Technologies in the Construction Industries of the United States and Korea." *Journal of Construction Engineering and Management* 133 (10): 780–790. doi:10.1061/(ASCE)0733-9364(2007)133:10(780).

Yoon, J. 2012. "Detecting Weak Signals for Long-Term Business Opportunities Using Text Mining of Web News." *Expert Systems with Applications* 39 (16): 12543–12550. Elsevier. doi:10.1016/j.eswa.2012.04.059.

Yoon, Byungun, Chang Byung Yoon, and Yong Tae Park. 2002. "On the Development and Application of a Self-Organizing Feature Map-Based Patent Map." *R&D Management* 32 (4): 291–300. doi:10.1111/1467-9310.00261.

Yu, Peng, and Jang Hee Lee. 2013. "A Hybrid Approach Using Two-Level SOM and Combined AHP Rating and AHP/DEA-AR Method for Selecting Optimal Promising Emerging Technology." *Expert Systems with Applications* 40 (1): 300–314. doi:10.1016/j.eswa.2012.07.043.

Zheng, Jia, Zhi-yun Zhao, Xu Zhang, Dar-zen Chen, and Mu-hsuan Huang. 2014. "International Collaboration Development in Nanotechnology: A Perspective of Patent Network Analysis." *Scientometrics* 98 (1): 683–702. Springer. doi:10.1007/s11192-013-1081-x.

Appendix 1. Visualising keyword intensity maps in the portfolio map structure

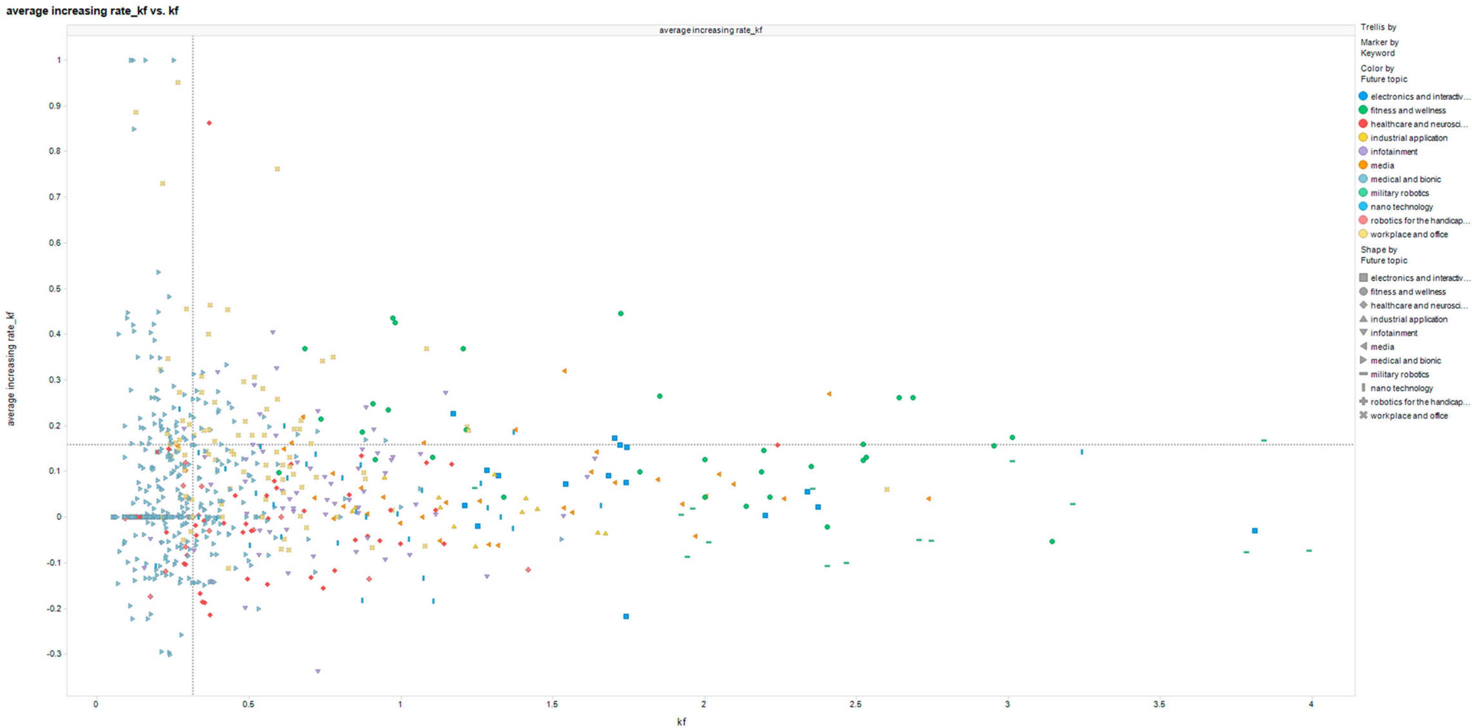

Figure A1. Keyword visibility map.

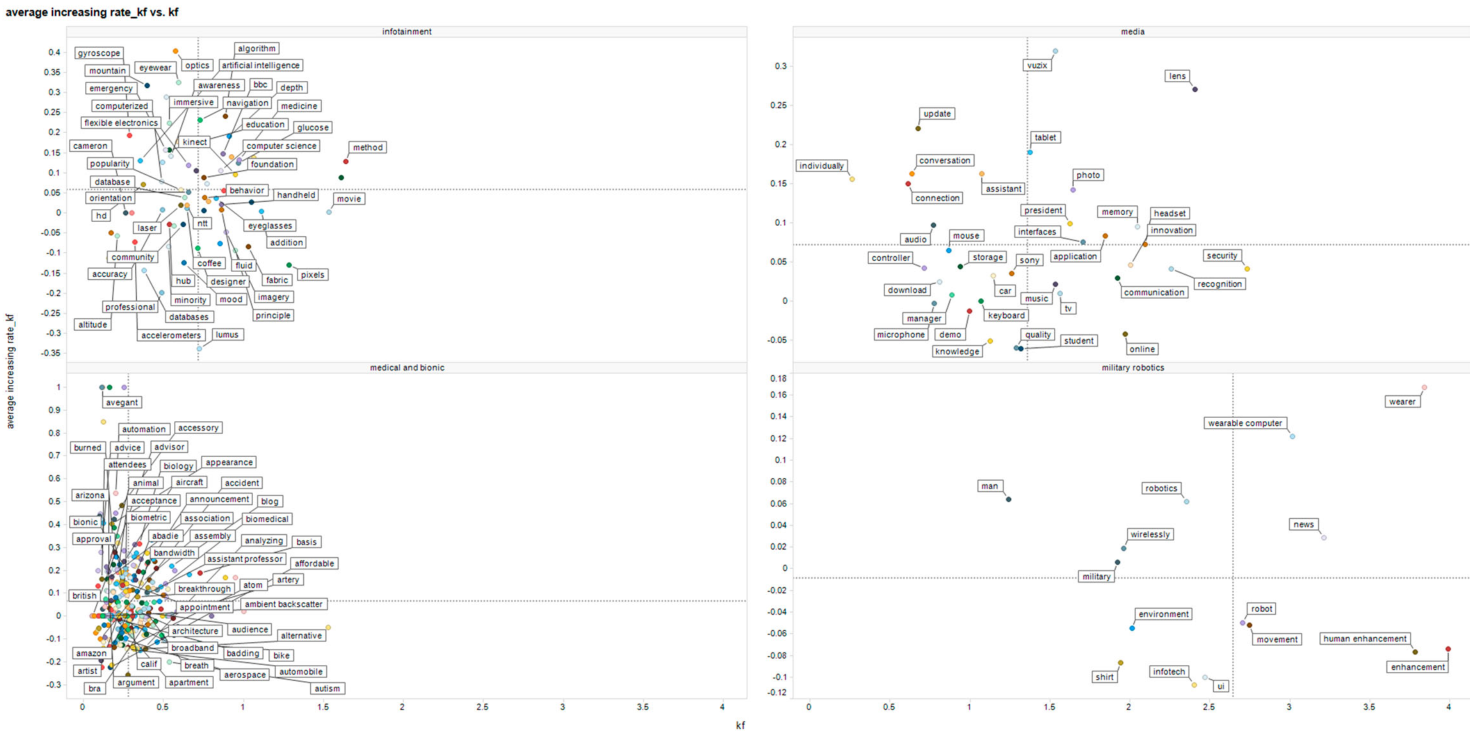

Figure A2. Keyword visibility map per topic (partial).

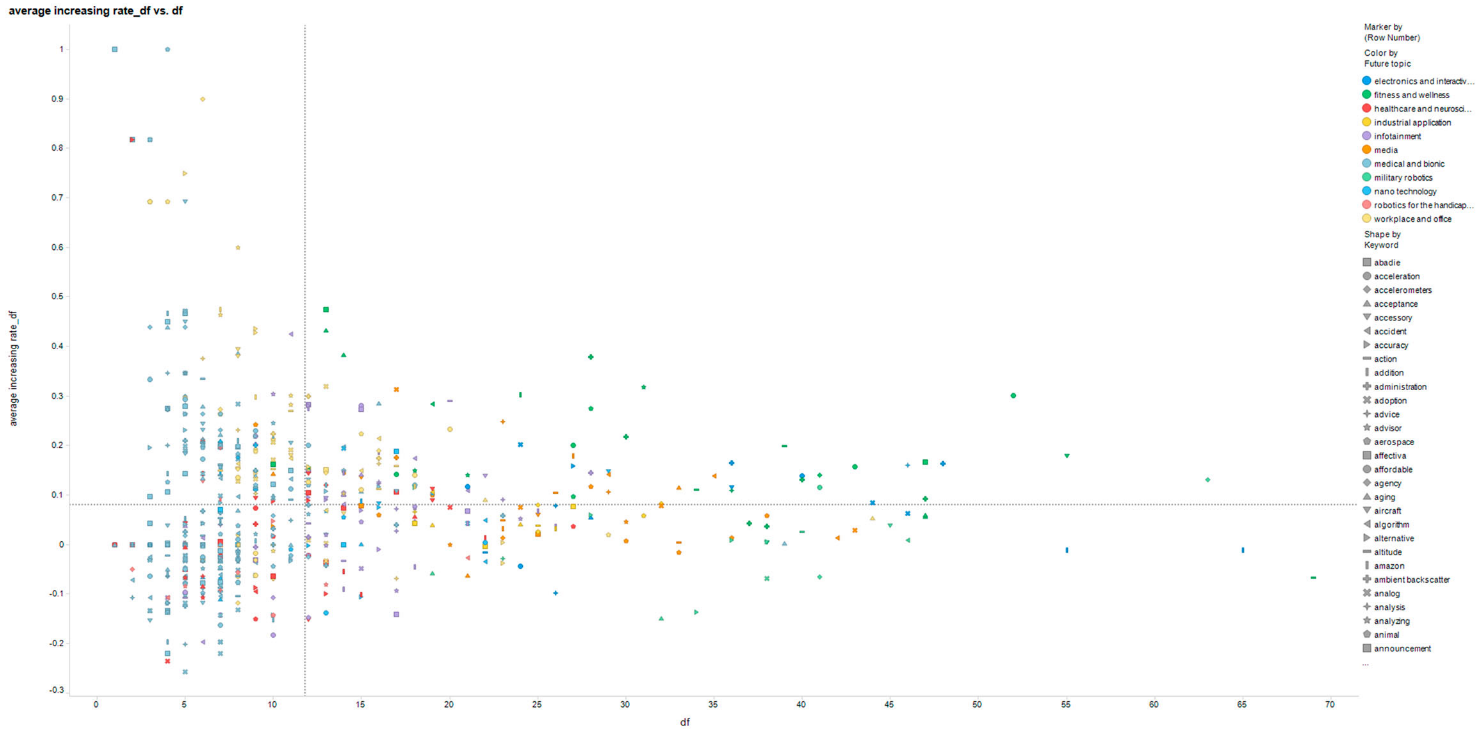

Figure A3. Keyword diffusion map.