Expert Systems With Applications 57 (2016) 311–323

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Futuristic data-driven scenario building: Incorporating text mining and fuzzy association rule mining into fuzzy cognitive map

Jieun Kim[a], Mintak Han[b], Youngjo Lee[c], Yongtae Park[b,*]

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

ARTICLE INFO

Article history:
Received 14 December 2015
Revised 24 March 2016
Accepted 25 March 2016
Available online 29 March 2016

Keywords:
Scenarios
Fuzzy cognitive map
Futuristic data
Text mining
Fuzzy association rule mining
Electric vehicle

ABSTRACT

Fuzzy cognitive maps (FCMs) are one of the representative techniques in developing scenarios that include future concepts and issues, as well as their causal relationships. The technique, initially dependent on deductive modeling of expert knowledge, suffered from inherent limitations of scope and subjectivity; though this lack has been partially addressed by the recent emergence of inductive modeling, the fact that inductive modeling uses a retrospective, historical data that often misses trend-breaking developments. Addressing this issue, the paper suggests the utilization of futuristic data, a collection of future-oriented opinions extracted from online communities of large participation, in scenario building. Because futuristic data is both large in scope and prospective in nature, we believe a methodology based on this particular data set addresses problems of subjectivity and myopia suffered by the previous modeling techniques. To this end, text mining (TM) and latent semantic analysis (LSA) algorithm are applied to extract scenario concepts from futuristic data in textual documents; and fuzzy association rule mining (FARM) technique is utilized to identify their causal weights based on if-then rules. To illustrate the utility of proposed approach, a case of electric vehicle is conducted. The suggested approach can improve the effectiveness and efficiency of scanning knowledge for scenario development.

1. Introduction

The uncertainty of the business environment has highlighted the strategic gravity of scenario in technology foresight and strategic planning (Bishop, Hines, & Collins, 2007). Scenarios are defined as a set of hypothetical events in the future constructed to clarify a possible chain of causal events as well as their decision points (Kahn & Wiener, 1967); or a disciplined methodology for imagining possible futures in which organizational decisions may be played out (Schoemaker, 1995). As the consideration of scenarios can significantly enhance the ability to deal with uncertainty and the usefulness of overall decision making process, scenario planning has been adopted in technology planning or strategic analysis (Drew, 2006; Hirsch, Burggraf, & Daheim, 2013).

Fuzzy cognitive map (FCM), among various scenario development approaches, has recently drawn attention due to its relative advantage in combining qualitative (creative) knowledge and quantitative structuring process (Amer, Daim, & Jetter, 2013a, 2013b; Biloslavo & Dolinšek, 2010; Jetter & Kok, 2014; Jetter & Schweinfort, 2011; Kok, 2009; Salmeron, Vidal, & Mena, 2012; Soler, Kok, Camara, & Veldkamp, 2012). FCMs are cognitive fuzzy inference graphs, within which the nodes stand for the concepts that are used to describe the behavior of the system and the causal relations between the concepts are represented by signed and weighted arcs (Kosko, 1986). Since the FCMs simulate dynamic evolution based on its initial model, they can be used to analyze and test the influence of parameters and predict the behavior of the system. Thus, FCM-based scenario approach is known to cover most of the generic set of steps for scenario planning (Jetter & Schweinfort, 2011)

The formalization of FCMs has been achieved by two main groups of methods: deductive modeling using expert knowledge about the domain of application, and inductive modeling using learning algorithms based on historical data. A number of previous studies applied FCM to model expert-based systems (Khan & Quaddus, 2004; Lee & Lee, 2015; Salmeron et al., 2012; Stach, Kurgan, & Pedrycz, 2010; Tsadiras & Bassiliades, 2013). Although these *deductive modeling* methods are well-established, they have shortcomings in that they require domain knowledge, which can be limited to relatively simple systems and subjective or biased

* Corresponding author. Tel.: +82 2 880 8358; fax: +82 2 889 8560.
E-mail addresses: hsns1234@snu.ac.kr (J. Kim), alsxkr77@snu.ac.kr (M. Han), youngjo@snu.ac.kr (Y. Lee), parkyt1@snu.ac.kr (Y. Park).

http://dx.doi.org/10.1016/j.eswa.2016.03.043
0957-4174/

models. To address such shortcomings, several *inductive modeling* methods have recently been proposed and examined to generate FCM models from historical input data without human intervention (Chen, Mazlack, Minai, & Lu, 2015; Papageorgiou, 2012; Stach, Kurgan, Pedrycz, & Reformat, 2005). Papageorgiou (2012) provides a Hebbian-based learning algorithm to produce weight matrices that lead the FCM to converge into a given decision state or an acceptable region, and Population based learning algorithm (e.g., genertic algorithm, particle swarm optimization, divide and conquer, etc.) to compute weight values on the basis of historical data that best fit the sequence of the states of concept nodes. However, the attempts of inductive modeling are also subject to fundamental limitations. First, the greater part of their focus is on the identification of weight values when the set of concept nodes are given by experts (Papageorgiou, 2012). Second, they rely on only historical data, a list of the phenomena regarding target system, and assume that same trends will prevail in the future.

In this context, we propose that the *futuristic data*, a collection of future-oriented opinions extracted from websites and online communities of large participation and collaboration of many experts and the general populace (Cachia, Compañó, & Costa, 2007; Markmann, von der Gracht, Keller, & Kroehl, 2012; Pang, 2010; Raford, 2015; Schatzmann, Schäfer, & Eichelbaum, 2013), can be supplementary or even alternative knowledge source for FCM-based scenario development. Recently, the emergence of information and communication technology (ICT) and Web 2.0 has enabled methodological innovation in foresight exercises including scenario planning. Raford (2015) explores the specific role that ICT may play in qualitative scenario planning. It is, in fact, found to have substantial impact on the early stages of the scenario process, including: increased participation in terms of both amount and diversity, increased volume and speed of data collected and analyzed, increased transparency around driver selection and analysis, and decreased overall cost of administration (Keller & von der Gracht, 2014).

As technology foresight websites are launched by many providers, meaningful and massive futuristic database are accumulating. In turn, a concept of future-related database has been suggested by a number of literatures. Schatzmann et al. (2013), for instance, collected existing digital collaborative prediction and foresight applications, and subsumed them into four categories: databases/wiki, prediction markets, social rating systems, and collaborative scenarios. Pang (2010) offered three strengths of online communities in technological forecasting: providing a platform to share the data, serving the evaluation of the output of forecasting and aiming to aggregate collective intelligence through online participation. Cachia et al. (2007) suggested that creativity of expert group derives from interactions and communications in online communities, because they can cover rapid changes and trends in social behaviors and responses. Markmann et al. (2012) analyze existing future-oriented database, so-called trend database, and identifies four major challenges of utilizing trend database, such as extensiveness, cooperation, linking, and incentive. Based on these concepts, this paper would like to analyze such futuristic data. Since the futuristic data are a priori, i.e., nonhistorical data containing issues and discussions for directions, expectations, and predictions of future, they are a suitable source from which to scan not only future drivers of changes and resulting impacts, which will be used as the concept nodes of FCM, but also the relationship among them, which will be used as the edges of FCM.

Taken together, the primary objective of this research is to propose the approach for applying futuristic data to FCM-based scenario development. Despite their utility, extracting the future drivers and their causal relationships from the vast amount of futuristic data can make scenario building more time-consuming (Mietzner & Reger, 2005). Thus, several data mining techniques are applied to the futuristic database to systematically identify patterns and develop the FCM. First, in order to identify the concept nodes of FCM, keywords and textual patterns are extracted from futuristic database using text mining (TM) (Berry & Kogan, 2010; Lin, Hsieh, & Chuang, 2009) and latent semantic analysis (LSA) (Dumais, 2004). Second, in order to identify the causal relationships and weights among concept nodes of FCM, fuzzy association rule mining (FARM) and Partial Association (PA) test are applied. Association Rule Mining (ARM) can provide if-then association rules from large database with high-dimensionality (Agrawal, Imieliński, & Swami, 1993). Unlike the traditional standard ARM, which requires the binary valued input data set, FARM can deal with a numeric attribute that can take a range of values; thus, it can consider the importance and frequency of concepts that appear in futuristic database. Furthermore, since the association rules do not directly indicate causal relationships, we utilize PA tests from Causal Rule-Partial Association (CR-PA) algorithm, suggested by Jin et al. (2012) and Li, Liu, and Le (2015), to exclude noncausal associations and to ensure the high reliability and persistence of causal rules. The suggested approach can presumably help improve the effectiveness and efficiency of scanning knowledge for FCM-based scenario development.

The rest of this paper is constructed as follows. Section 2 reviews the methodological backgrounds of FCM-based scenario building, TM and LSA, and FARM. Section 3 proposes the approach to futuristic data-driven scenario building. Section 4 illustrates the feasibility and utility of proposed approach from the case study of Electronic Vehicle. Finally, the paper ends with Section 5 including discussions and conclusions.

2. Theoretical background

This section reviews the detailed theory and method of previous FCM-based scenario approach, TM and LSA, and FARM, which will be utilized and integrated for futuristic-data driven scenario building.

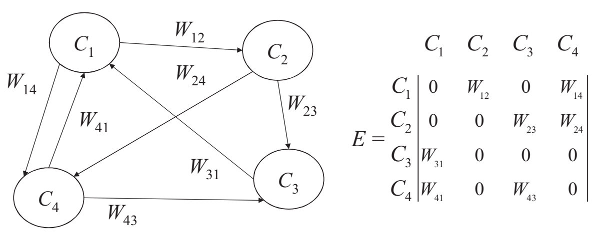
2.1. Fuzzy cognitive map (FCM)-based scenario approach

This paper aims to improve previous FCM-based scenario approach by integrating several approaches, we firstly investigate FCM-based scenario approach. Since FCM is general method utilized in various application areas, we review basic fundamentals of FCM and its utilization for scenario planning.

2.1.1. Fundamentals of FCM

FCM, firstly suggested by Kosko (1986), is the extension and enhancement of cognitive map to present a belief system in a given domain and is developed by experts using interactive procedure of knowledge acquisition (Yaman & Polat, 2009). As the name suggests, FCMs originated from a combination of fuzzy logic and neural networks (Motlagh, Jamaludin, Tang, & Khaksar, 2015; Papageorgiou, 2012) and describe the behavior of a system in terms of concepts and their causal relationships. In FCMs, the nodes stand for the *concepts* that are used to describe the behavior of the system (e.g., an entity, a state, a variable, or a characteristic of the system) and the causal relations between the concepts are represented by *signed and weighted arcs*. The detailed elements are as follows:

- *Concepts:* C_1, C_2, ..., C_n. These represent the drivers and constraints that are considered of importance to the issue under consideration.
- *State vector:* $A = (A_1, A_2, ..., A_n)$, where a_i denotes the state of the node C_i. The state vector represents the value of the concepts, usually between 0 and 1. The dynamics of the state vector A is the principal output of applying a FCM.



(a) FCM that consists of concepts $C_1, \ldots, C_4$ and directed edges with fuzzy weights W_{12}, W_{14}, etc.

(b) Adjacency matrix (E) that contains the fuzzy weights of the directed edges of (a)

Fig. 1. Structure of a FCM and the corresponding adjacency matrix.

- *Directed edges:* $C_1 \to C_2$, etc. These represent the relationships between concepts, visualized as arrows in the directed graph.
- *Adjacency matrix:* $E = (W_{ij})$, where W_{ij} is the fuzzy weight of the directed edge C_i to C_j. The matrix contains the values of all relationships between concepts, usually between -1 and $+1$. Note that contrary to most applications, non-zero values on the diagonal are considered here.

Fig. 1(a) shows the graphical representation of an FCM that consists of four concepts $C_1, \ldots, C_4$, and seven directed edges with fuzzy weights W_{12}, W_{14}, etc. Each concept node has its own state value between 0 and 1. The weights of edge are between -1 and $+1$ and configure the adjacency matrix as shown in Fig. 1(b). The visual illustration of FCM can show the interconnections between concepts and facilitate suggestions in the reconstruction of the FCM, as the adding or deleting of interconnection of a concept.

The strength of FCMs in dynamic modeling comes from *inference mechanism*. The causality relationships of a FCM model are dynamic and cumulative and thus have a time dimension. Given the FCM with a number of nodes C_i (where $i = 1, \ldots, n$), the value of each node in iteration can be updated as:

$$A_i(t+1) = S(A_i(t) + \sum_{\substack{j \neq i \\ j=1}}^{N} A_i(t) \cdot W_{ji}) \tag{1}$$

where $A_i(t+1)$ is the value of concept C_i at the time $(t+1)$, $A_i(t)$ is the value of concept C_i at the time t, W_{ij} is the fuzzy weight of the directed edge C_i to C_j, and S is a threshold (activation) function that squashes the result of the multiplication in the interval [0, 1] wherein concepts take values. The nonlinear function S can be of various types such as bivalent, trivalent, logistic, and sigmoid (Tsadiras, 2008). Based on these concepts, the inference mechanism is implemented as follows (Yaman & Polat, 2009): (1) the FCM is initialized and the activation level of each node for threshold function is set; (2) the node interaction is allowed by repeated matrix multiplication between state vector (A) and the weight of edges (W_{ij}); and (3) this interaction continues until stabilization, limit cycle, or chaotic behavior (Tsadiras, 2008). In stabilization case, the state values of concept nodes fluctuate in the early iteration (transient-state) but stabilize as fixed-point equilibrium is reached (steady-state).

FCMs have several advantages in that they are simple and intuitive, easy to understand their formalization as well as execution. Thanks to its flexibility in representation (as more concepts/phenomena can be included and linked), individual FCMs from different experts and/or stakeholders can be easily merged (Khan & Quaddus, 2004; Stach et al., 2010). Thus, FCMs can effectively model qualitative knowledge into quantitative structuring and analyzing processes (Amer & Daim, 2013; Jetter & Kok, 2014; Kok, 2009). They can consider not only the activation function like other neural network systems but also weight training to learn about relationships among contributing factors (Motlagh et al., 2015). Thanks to these advantages, FCMs have been widely used to predict the outcome of interactions between concepts, i.e., to perform *what-if experiments* (Amer, Daim, & Jetter, 2013b). The domain for their application is vast, including business and management, engineering, computer science, chemistry, medicine, environment and ecology, education, decision sciences, etc. (Papageorgiou & Salmeron, 2013).

2.1.2. FCMs for scenario planning

Many previous studies have focused on application of FCM into future scenario building (Amer et al., 2013b; Biloslavo & Dolinšek, 2010; Ferreira, Jalali, & Ferreira, 2015; Jetter & Kok, 2014; Jetter & Schweinfort, 2011; Salmeron et al., 2012). The main purpose of scenario planning is to focus on the uncertain aspects of the future and develop a limited number of possible states that tell a story of how various elements might interact under certain conditions (Schoemaker, 1995). For this, FCM-based scenario approaches elicit diverse experts' knowledge of uncertain driving forces that shape the future and simulate what-if experiments to create the alternative scenarios (Jetter & Kok, 2014). The following framework is a typical process for development of FCM-based scenarios by integrating scenario planning and FCM modeling processes (Jetter & Schweinfort, 2011):

- *Step 1: Scenario preparation:* Clarification of the objective, time frame and boundaries of the scenario project.
- *Step 2: Knowledge capture:* Identify relevant concepts/potential scenario drivers through experts and literature review, merge mental models of various experts and subsequently translate these into conceptual FCM scenario model.
- *Step 3: Scenario modeling:* Streamline the causal links and assign weights and signs to all links, choose threshold functions for all concepts.
- *Step 4: Scenario development:* Calculate the FCM model for different *input vectors* that represent plausible combinations of concept states.
- *Step 5: Scenario selection and refinement:* These raw scenarios developed after step 4 are further assessed and refined by scenario simulation.
- *Step 6: Strategic decisions:* The developed scenarios are used for making strategic decisions.

In step 4, in order to construct the input vectors, uncertain and important concepts as well as their plausible combinations should be identified. Several previous studies have used static analysis to identify important concepts based on network theory such as centrality (Ferreira et al., 2015; Khan & Quaddus, 2004; Yaman & Polat, 2009), and morphology analysis (Ritchey, 2006; Yoon, Phaal, & Probert, 2008) to investigate raw scenarios, which is the combination of important concepts (Amer et al., 2013b). On the other hand, in step 5, dynamic analysis is conducted to identify the final results of the simulation of the input vectors of the raw scenarios, using the above-mentioned inference mechanism (Amer et al., 2013b).

2.2. Text mining (TM) and latent semantic analysis (LSA)

TM, which covers the process of finding interesting patterns, models, directions, trends, or rules from unstructured text, is an automated discovery of knowledge from texts (Berry & Kogan, 2010; Lin, Chen, & Tzeng, 2009). Structuring the input text usually involves parsing, along with the addition and removal of derived linguistic features, and subsequent insertion into a database. TM assumes that documents in the text format can be featured by keywords and thus a term-document matrix is the general method of handling large amounts of unstructured text to extract information from structured data (Lin et al., 2009).

Table 1
Measures of ARM.

Measure	Formula	Meaning and implication
Support	$supp(X \to Y) = P(X \cap Y) = \frac{N(X \cap Y)}{N}$	The usefulness of discovered rules
Confidence	$conf(X \to Y) = P(Y\|X) = \frac{P(X \cap Y)}{P(X)}$	The certainty of the rule
Lift	$lift(X \to Y) = \frac{conf(X \to Y)}{\sup p(Y)} = \frac{P(X \cap Y)}{P(X)P(Y)}$	The statistical dependence between items X and Y If the lift value is greater than one, it shows a positive correlation

LSA is a technique that analyzes relationships between a set of documents and the terms they contain by producing a set of concepts related to the documents and terms (Dumais, 2004; Landauer, Foltz, & Laham, 1998). It assumes that terms that are close in meaning will occur in similar pieces of text. By lowering the dimension of term-document matrix by Singular Value Decomposition, it is able to distinguish multiple words that have similar meanings and words that have more than one meaning. Also, the significance of each semantic textual pattern is figured out through LSA. LSA has been used typically in information retrieval and clustering or classification of documents or terms (Dumais, 2004). Recently, some pioneering works have been introduced to apply LSA to technology foresight area such as classifying applied science research projects (Thorleuchter & Van den Poel, 2013), weak signal tracing (Thorleuchter, Scheja, & Van den Poel, 2014), cross impact analysis (Thorleuchter & Van den Poel, 2014), and technology opportunity analysis (Zhu & Porter, 2002).

2.3. Fuzzy association rule mining (FARM)

Association Rule Mining (ARM) is a popular and well researched method for discovering interesting relationships between variables in large databases. Typical application of ARM is market basket analysis for identifying a set of product items frequently purchased together in large-scale customer transaction data (Agrawal et al., 1993). ARM provides information of correlations, frequent patterns, associations or casual structures among sets of items in the form of 'if-then' rules, e.g., {Diapers}→{Beer}. Formally, let $I = \{i_1, i_2, \dots, i_m\}$ be a set of items, a transaction database $D = \{t_1, t_2, \dots, t_N\}$ is set of transactions, where each transaction is set of items such that $t_j \subseteq I$. A rule is expressed as $X \to Y$, where X, $Y \subseteq I$ and $X \cap Y = \phi$, which means that if the transaction includes item X ("antecedent"), then it also includes item Y ("consequent"). If $P(X)$ is the probability that item X is included in transaction t_j such that $X \subseteq t_j$, $N(X)$ is denoted as the count of transactions containing item X, and N is the total number of transactions in T, ARM includes three core measures to generate and select association rules as represented in Table 1.

Based on these measures, the typical procedure of ARM consists of two steps (Agrawal et al., 1993): (1) generate frequent itemset – to create all item combinations greater than or equal to a minimum support (*minsupp*) threshold, (2) identify association rules – to select itemsets greater than or equal to a minimum confidence (*minconf*) threshold among the frequent itemsets found in (1). There are several techniques for step (1) and the most representative one is Apriori algorithm.

One fundamental limitation of classical standard ARM is that all attributes of I are required to be binary valued; in transaction database, the values are whether the transaction contains the item or not (Agrawal et al., 1993). Thus, to handle databases with both categorical and quantitative attributes, a quantitative association rule mining method was proposed by (Srikant & Agrawal, 1996). The method finds association rules by partitioning the quantitative attribute domain and then transforming the problem into binary one. Apparently, whatever partitioning methods are applied, "sharp (crisp) boundaries" remain a problem, which may lead to an inaccurate representation of semantics. As a remedy to the sharp boundary problem, the fuzzy set concept has recently been used more frequently in mining quantitative association rules as FARM (Farzanyar & Kangavari, 2012).

The principal idea of FARM is that ranged values can belong to more than one sub-range, thus the value has a membership degree that associates it with each available sub-ranges. Using previous notations, let $I = \{i_1, i_2, \dots, i_m\}$ be the item set where each i_j is an attribute of the original dataset, $D = \{t_1, t_2, \dots, t_N\}$. Although each attribute i_j was binary in general ARM, i_j here may have a binary, categorical, or quantitative underlying domain Δ_j. Besides, each item i_j is associated with its fuzzy sets, extending the item set I_f from I. Using the corresponding membership functions defined with each fuzzy set, the original dataset D is changed into a fuzzy dataset D_f. Given the fuzzy dataset $D_f = \{t_1, t_2, \dots, t_N\}$ with I_f, the discovered rules are of the form same as standard ARM, $X \to Y$, where t_i is a transaction in D_f, X, $Y \subseteq I_f$ and $X \cap Y = \phi$. In this setting, the definitions of support and confidence measures of the rule $X \to Y$ for the whole D_f are extended as (Farzanyar & Kangavari, 2012):

$$sup\ p(X \to Y) = \frac{\sum_{i=1}^{N} X(a) \otimes Y(b)}{\left|D_f\right|} \tag{2}$$

$$conf(X \to Y) = \frac{\sum_{i=1}^{N} X(a) \otimes Y(b)}{X(a)} \tag{3}$$

where $|D_f|$ is the total number of transactions in D_f, which is equal to N, the number of transactions in the quantitative database D. $X(a)$ and $Y(b)$ are the membership degree of the elements a and b with respect to the fuzzy set X and Y respectively, $\otimes$ denotes a 't-norm' that aggregates the intersection of two membership degrees. Based on the notations, the procedure to extracting rules follows same logic using *minsupp* and *minconf* defined by users.

3. Futuristic data-driven scenario building

3.1. Outline of proposed approach

The core idea of this paper is the application of TM and FARM to FCM in order to leverage the textual futuristic data into scenario building. To this end, the proposed method consists of five stages, as shown in Fig. 2.

First, futuristic data are collected from the technology foresight websites. Second, a set of terms are extracted and structured as the term-document matrix and scenario concepts are constructed, by applying TM and LSA to collected data. The '*scenario concept*' in this paper is technically defined as a topic which consists of the set of keywords with similar meaning; LSA is applied for identifying the coherent set of keywords (i.e., semantic textual patterns) and defining scenario concepts. Third, the causal relationships and weights are identified as the result of FARM. Fourth, a scenario model is finalized and visualized as a FCM. Lastly, the FCM-based scenario is analyzed in static and dynamic viewpoint.

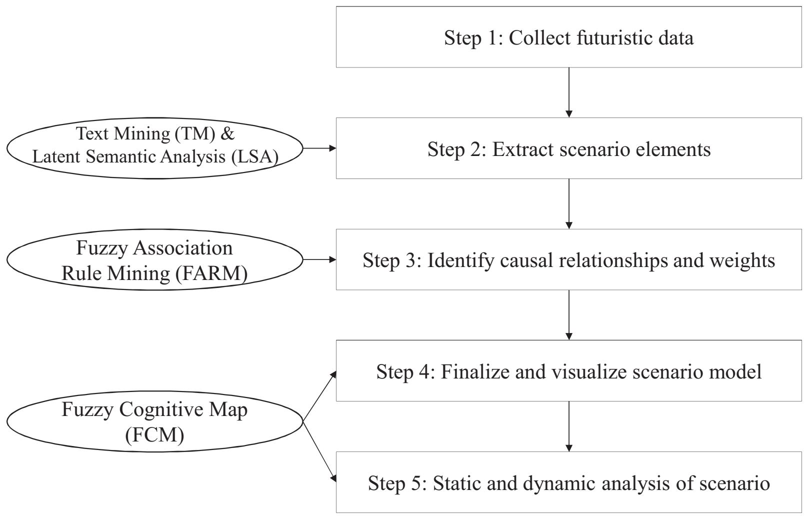


Fig. 2. Overall research process of futuristic data-driven scenario building.

Table 2
Examples of futuristic data (adapted and modified from Schatzmann et al., 2013).

Types	Description	Example of source websites
Databases and Wikis	A digital archive that provides a classification schema for future-related information such as wildcard databases, prediction databases and reports, trend databases and reports, databases that are used for horizon scanning and databases that are used for mapping strategic foresight; Wikis are one type of database that especially provide the collective knowledge and ontologies of information	Siemens, IBM IT Insight, GE-Technologist, GE-Ideas Lab, iKnow, TechCast, TrendWiki
News and blogs	News and blog posts including replies that offer technological trends and predictions of the individual and their interaction in online websites such as expert forums, trend review websites, etc.	Gartner, McKinsey, MIT Technology Review, Kurzweil Accelerating Intelligence, Next Big Future, World Future Society, LongBets, iKnow, Trendradar2020,Shaping Tomorrow, SigmaScan, Delta Scan, Forecasting World Events, The Seven Horizons, wrong tomorrow, Future Scanner, TechCast, SKAN, Vanguard Science Daily, Engadget, GSM Arena, TechCrunch, The Verge
Social rating systems	The vast set of assumptions, predictions, conjectures, and the rates of them using the scales like relevance, impact, likelihoodor desirability	Is it Future proof?, bean sight, The Future of Facebook Project, predicto.net, wefutr, Web of Fate, Wikistrat, The Future of Facebook Project, Forecasting ACE, NY Times Technology Timeline, Real-time Delphi
Collaborative scenarios	The data generated from collaborative scenarios, i.e., the aggregated descriptions on future predictions made by many participants; The focus is on pooling only potential scenarios and solutions to specific future challenges and the ability to aggregate assumptions about the future into scenarios	superstruct, signtific map, Risk Interconnection Map, Future Timeline, News of Future
Prediction markets	The data generated from prediction markets, i.e., exchange-traded markets created for the purpose of trading the outcome of events; The payoff of contracts depends on the possible events in the future.	intrade, inklingmarket, Popular Science Predictions, Betfair, Iowa Electronic Markets, Smarkets, iPredict, Predictious, Prediction Lab, PredictIt, SciCast, Hypermind

3.2. Detailed process

3.2.1. Collecting data

Futuristic data is defined as a collection of future-oriented opinions extracted from websites and online communities of large participation and collaboration of many experts and the general (Kim & Park, 2014). The examples of futuristic data contained in such websites are listed in Table 2: database and wikis, news and blogs, social rating systems, collaborative scenarios, and prediction markets. The providers of futuristic data can be various stakeholders such as IBM and GE, professional technology forecasting or consulting companies such as Gartner and McKinsey, trend reporting websites such as Science Daily and Engadget, communities of experts or futurists such as Next Big Future, World Future Society, and communities of the public such as social rating system, collaborative scenarios, and prediction markets. The field of technology they focused is mainly ICTs, but other fields such as bio, nano, and energy, and social trends are also part of their concerns. Futuristic data include various level of future-oriented information such as current trend, short-term forecasts, or long-term forecasts. Also the forms of information are various, such as news, report, magazine, web post (blog), forum (tread-reply), etc.

From these various sources of futuristic data, a scenario developer should identify which websites are offering the information regarding future scenario he/she want to build. Since the scenarios can also focus on various technology fields (e.g., IT scenario, BT scenario, etc.) or level of future-orientation (e.g., short-term scenario or long-term scenario), the source websites can be selected by considering the information characteristics of websites such as technology field and future-orientation and data volume of searching targeted scenario subjects.

3.2.2. Extracting scenario concepts

This paper identifies scenario concepts (i.e., concepts that configure a FCM) by extracting a set of keywords that have semantic similarity. Since the futuristic data are textual, unstructured corpora (i.e., set of documents), TM is necessary to process their natural languages. TM starts with extracting terms from corpora. Since there are meaningless keywords such as stopwords, or the keywords with low frequency, the overall keyword set is refined by eliminating such keywords. Then, the term-document matrix (TDM) with the normalized frequency of keywords is constructed, as below:

$$TDM = \begin{array}{c} \\ t_1 \\ t_2 \\ \dots \\ t_n \end{array} \begin{array}{c} \begin{array}{cccc} d_1 & d_2 & \dots & d_m \end{array} \\ \left| \begin{array}{cccc} tf_{11} & tf_{12} & \dots & tf_{1m} \\ tf_{21} & tf_{22} & \dots & tf_{2m} \\ \dots & \dots & \dots & \dots \\ tf_{n1} & tf_{n2} & \dots & tf_{nm} \end{array} \right| \end{array} \tag{4}$$

where t_i is ith term extracted from corpora (where $i = 1,\dots, n$), d_j is jth document (where $j = 1,\dots, m$), and tf_{ij} is the normalized term frequency of t_i in document d_j. However, t_i, keywords in TDM are not enough to grasp the context of textual data and analyze their latent meanings. The reason is that the futuristic data are online documents freely written by many users who can have different writing styles or use different words for same meaning. In this case, LSA is effective to identify the set of keywords with similar meanings, which delivers one topic of documents. Thus, as we defined previously, the scenario concepts are defined by semantic textual patterns that LSA derives as the set of term dimensions that can be integrated as one dimension in TDM. To do this, Singular Value Decomposition is performed on TDM to reduce the dimension of the matrix (Dumais, 2004) as follows:

$$TDM = U\Sigma V^T$$

$$U = \begin{array}{c} \\ t_1 \\ t_2 \\ \dots \\ t_n \end{array} \begin{array}{c} \begin{array}{cccc} v_1 & v_2 & \dots & v_p \end{array} \\ \left| \begin{array}{cccc} u_{11} & u_{12} & \dots & u_{1p} \\ u_{21} & u_{22} & \dots & u_{2p} \\ \dots & \dots & \dots & \dots \\ u_{n1} & u_{n2} & \dots & u_{np} \end{array} \right| \end{array} \quad \Sigma = \begin{array}{c} \\ v_1 \\ v_2 \\ \dots \\ v_p \end{array} \begin{array}{c} \begin{array}{cccc} v_1 & v_2 & \dots & v_p \end{array} \\ \left| \begin{array}{cccc} \Sigma_1 & 0 & \dots & 0 \\ 0 & \Sigma_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \Sigma_p \end{array} \right| \end{array} \quad V = \begin{array}{c} \\ d_1 \\ d_2 \\ \dots \\ d_m \end{array} \begin{array}{c} \begin{array}{cccc} v_1 & v_2 & \dots & v_p \end{array} \\ \left| \begin{array}{cccc} v_{11} & v_{12} & \dots & v_{1p} \\ v_{21} & v_{22} & \dots & v_{2p} \\ \dots & \dots & \dots & \dots \\ v_{m1} & v_{m2} & \dots & v_{mp} \end{array} \right| \end{array} \tag{5}$$

where u_{ik} is the impact of ith term on kth semantic textual pattern for U (where $k = 1,\dots, p$), $n \times p$ matrix with orthonormal columns, Σ_k is the importance of kth semantic textual pattern for Σ, $p \times p$ diagonal matrix with the entries sorted in decreasing order, and v_{jk} is the impact of jth document on kth semantic textual pattern for V, $m \times p$ matrix with orthonormal columns. Thus, the semantic textual patterns v_1, v_2, ..., v_p are directly mapped into scenario concepts, C_1, C_2, ..., C_p in FCM. For each v_k, the terms highly impacting v_k are identified by investigating u_{ik} in the matrix U, and corresponding scenario concept C_k is defined considering the meaning of allocated terms with high u_{ik}.

3.2.3. Identifying causal relationships and weights

In order to extract association rules for the scenario concepts, C_1, C_2, ..., C_p, the matrix V, the impact of documents on concepts is used as the input of FARM. The reason why we apply FARM is to extract causal rules among concepts such as $C_k \rightarrow C_l$, considering the quantitative impact values of concepts in documents. As described in Fig. 3, the fuzzy item set I_f here corresponds to set of scenario concepts $C = \{C_1, C_2, \dots, C_p\}$, whereas the fuzzy transaction data set $D_f = \{t_1, t_2, \dots, t_N\}$ here can be a set of documents containing the strength or impact of occurrence of C. Since the impact values $v_{jk}$of V matrix denotes the degree that document d_j impacts on C_k by containing the concept C_k in them, we interpret this as the degree of membership to be C_k.

The support, confidence, and lift measures are modified as follows:

$$\sup p(C_k \rightarrow C_l) = \frac{\sum_{j=1}^{m} v_{jk} \otimes v_{jl}}{m} \tag{6}$$

$$conf(C_k \rightarrow C_l) = \frac{\sum_{j=1}^{m} v_{jk} \otimes v_{jl}}{v_{jk}} \tag{7}$$

$$lift(C_k \rightarrow C_l) = \frac{conf(C_k \rightarrow C_l)}{\sup p(C_l)} \tag{8}$$

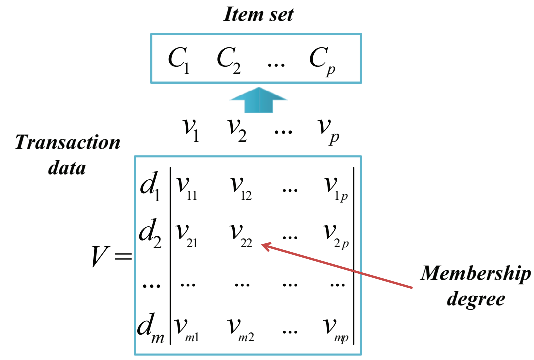


Fig. 3. Structure of input data for FARM, where fuzzy item set is a set of scenario concepts C_k defined by semantic textual patterns v_k, fuzzy transaction data set is a set of impact values v_{jk} of V matrix which denotes membership degree of document d_j to impact on C_k.

The output rules of FARM are represented as $C_k \rightarrow C_l$. These extracted association rules, however, do not directly indicate causal relationships. Causal relationships imply associations but reverse is not always true. Thus, the association rules are considered as causal hypotheses (i.e., candidate causal relationships) and reassessed whether they are causal rule or not. To elaborate this, we utilize partial association (PA) tests (Jin et al., 2012; Li et al., 2015). For association rule $X \rightarrow Y$, it assesses whether the rule is still associated when a third control variable Z is given. When both X and Y are caused by Z; or X causes Z and Z causes Y: in either case, X is not directly cause of Y thus the rule is not validated (i.e., zero partial association). The Mantel–Haenszel (MH) test (Mantel & Haenszel, 1959) is commonly used for testing the null hypothesis of zero partial association between two variables X and Y in any of the strata of a population, against the alternative that the degree or strength of partial association is nonzero. In our case, the association rule $C_k \rightarrow C_l$ and a set of control variables are remaining scenario concepts C_1, C_2, ..., C_p except C_k and C_l. Since the inputs for MH statistics are binary variables, V matrix is transformed as binary values using average as threshold. Then, the rule is tested for every value of control variables. The detailed definition of test statistic and description is appeared in the Li et al.'s (2015) study (p. 35). Since the MH test statistic has a Chi-square distribution, given a significance level α, if $MH \geq \chi_{\alpha}^{2}$, the null hypothesis of independence is rejected, and consider that the partial association between C_k and C_l is significant. So we conclude that the significant association rule is a causal rule, which will be used as one directed edge of FCM.

3.2.4. Finalizing and visualizing scenario model

As shown in Fig. 4, this study utilizes the support measure to eliciting significant concepts from the set of concepts C_1, C_2, ..., C_p, by maintaining concepts with *minsupp*. If every rules engaged with a concept has support values less than *minsupp*, the concept is filtered out from FCM. Likewise, the confidence values are also used for identifying association rules that satisfies *minconf*. Furthermore, the confidence values are utilized as the weight of edges, which corresponds to the 'absolute' values of W_{ij}. Since the confidence is

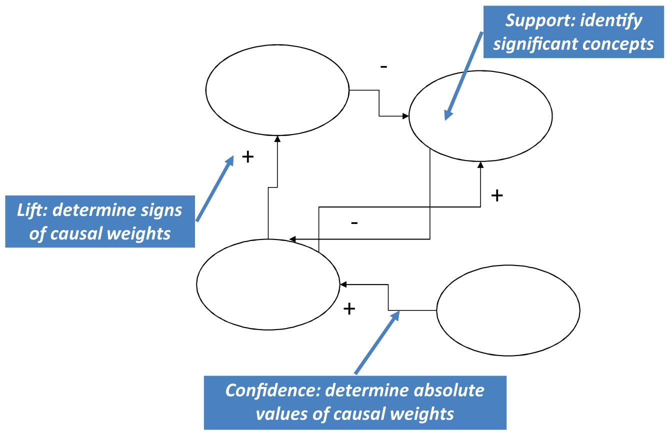


Fig. 4. Concept of applying FARM to FCM modeling, where concept nodes are determined by support, absolute values of causal weights are determined by confidence, and positive/negative signs are determined by lift.

the conditional probability of rules, or causal relationships, it directly corresponds to the impact of occurrence of one concept to another or the relative strength of relationships. Lastly, the signs of weights are determined by lift measure. Lift (Brin, Motwani, Ullman, & Tsur, 1997) refers to the ratio of the rule's confidence to the consequent's occurrence probability of the rule, representing how many times more often the antecedent and consequent occur together expected if they were statistically independent. Thus, if a lift $(X \to Y)$ is 1, it would imply that X and Y are independent each other. Moreover, lift is known to reflect the positive or negative correlation between the antecedent and consequent (Yongmei & Fuguang, 2015). If lift $(X \to Y) < 1$, then X and Y appear less frequently together in the data than expected under the assumption of conditional independence: they are said to be negatively interdependent. Likewise, if lift $(X \to Y) > 1$, then X and Y appear more frequently together in the data than expected under the assumption of conditional independence: X and Y are said to be positively interdependent. Using this framework, the signs of extracted rules can be determined.

3.2.5. Analyzing scenarios

For the constructed FCM, two types of analysis can be conducted. First, an FCM can be used for a *static analysis* of the domain for establishing the relative importance of concepts, and indirect and total causal effects between concept nodes (Khan & Quaddus, 2004; Yaman & Polat, 2009). The centrality of network theory can be a measure for determining the importance of nodes in an FCM, as follows:

$$Concept\ centrality = IN(C_i) + OUT(C_i) \tag{9}$$

where indegree, $IN(C_i)$ is the sum of the weights of causal links constituting all path connecting nodes C_j, $i \neq j$, to C_i and outdegree, $OUT(C_i)$ is the sum of weights of causal links constituting all path connecting node C_i to all nodes C_j, $i \neq j$. Concepts with high centrality values deserve special attention in any analysis for decision support.

Second, an FCM can be used for a *dynamic analysis* to observe and explore the impact of changes and the behavior of system with time (Amer et al., 2013b; Khan & Quaddus, 2004; Lee & Lee, 2015). Using dynamic analysis, a range of what-if analyses can be done by subjecting the FCM to a range of initial state vector values of interest. The important factors with high centrality can be utilized. According to Amer et al. (2013b), the input vectors of raw scenarios are constructed by a morphology analysis that combines and selects variations of important concepts. Then, the output vectors of several raw scenarios are identified by repeated multiplication of input vector and adjacency matrix until stabilization as dicussed in FCM inference mechanism.

4. Illustrative case study: electric vehicle (EV) scenarios

4.1. Background

In order to illustrate the applicability of the proposed approach, a case study of electric vehicle (EV) is conducted. EV is a collection of diverse technologies such as battery, motor, and recharging system. With the recent widespread deployment of EV technologies, urban mobility is expected to become more environmentally sustainable. In addition to technological aspects, however, there are a lot of social, economic, and political factors acting in a highly complex manner. Thus, a notable change may occur by one of the factors in the future development of EV.

There have been some studies to comprehend or assess the impact of EV in the near future (Lin et al., 2009; Zhang, Guo, Wang, Zhu, & Porter, 2013). For example, Lin et al. (2009) applied the DEMATEL technique, which is a comprehensive method for building and analyzing a structural model involving causal relationships between the complex factors to construct a cognition map of alternative fuel vehicles. Meanwhile, Zhang et al. (2013) generated global EV technology roadmap using hybrid model that combines bibliometric approaches, and experts' knowledge. However, the presented results mainly depend on few experts; and are lack of consideration of the collective intelligence on various factors. In this illustration, we consider a lot of opinions and predictions about the future of EV which have been discussed in futuristic database.

4.2. Process and results

4.2.1. Collecting data

As the data sources for futuristic documents, five websites are selected as Siemens (http://www.siemens.com/innovation/en/publications/index.htm), MIT technology review (http://www.technologyreview.com/topics/), Kurzweil Accelerating Intelligence (http://www.kurzweilai.net), World Future Society (http://www.wfs.org), and FutureTimeLine (http://futuretimeline.net/index.htm). The website of Siemens, a German multinational engineering and electronics conglomerate company, provides the future magazine reports that predicted technologies capable of changing the daily after 10–20 years. The reports describe R&D-related future scenarios along with the image and interviews with the experts who have been worldwide attention. Other websites including MIT technology review, Kurzweil Accelerating Intelligence, World Future Society, and FutureTimeLine – the communities of future experts or futurists – offer a number of articles (i.e., web posts, blogs, or news) regarding how various products and technologies that have attracted attention recently (e.g., smart watch, electric car, or iPad) will be changed and which elements of the product are future-oriented. The World Future Society is the largest nonprofit educational and scientific organization in the futures field, and they share future trends and perspectives in their websites. FutureTimeLine has a speculative timeline of the future history (from 21st century to beyond 1 million AD), which is based on the collaboration of scientists, futurists, inventors, writers and anyone else interested in futurology. As of March 2015, we collect documents related to future of EV from five foresight communities and the total corpora are constructed as 2017 futuristic documents.

4.2.2. Extracting scenario concepts

First, the futuristic data are processed by Natural Language ToolKits (NLTK), which is widely used in performing the Natural Language Process (NLP) for text data. This means that the step

Table 3
Part of the term-document matrix.

	Doc 1	Doc 2	Doc 3	Doc 4	…	Doc 2017
Acid battery	3	1	0	0		3
Atmosphere	1	3	2	0		0
Biofuel	0	8	1	0		0
…						
Vibration	0	0	0	1		2

is pre-process in text mining to determine the part of speech of each word, and remove the stop words such as articles, prepositions, and conjunctions. Then, the overall keyword vector is refined by eliminating the keywords with relatively low frequency. Based on the obtained keyword vector, the term-document matrix (TDM) with the frequency of keywords in each document is constructed. As a result, TDM consists of 92 keywords with 2017 documents. Part of the matrix is shown in Table 3. After that, normalized frequency of keywords with respect to the length of document is used as a value because frequency depends on the length of document.

For the next step, to figure out latent meaning of each term in the context of textual data, LSA is conducted with Matlab code. The reduced matrix by performing Singular Value Decomposition provides information on whether a higher impact of term in a particular semantic textual pattern. Thus, each semantic textual pattern can be defined as a concept in FCM considering the meaning of allocated terms with high impact value. From the data in Table 4, for example, it is apparent that the first semantic textual pattern is related to tourism. Consequently, the first concept is defined as 'application to tourism' in accordance with the top five terms with high value of impact. Total 15 concepts of FCM derived through the same procedure as above are presented in Table 4.

For the preliminary understanding of scenario concepts, we classified 15 concepts into STEEP (S: Social, T: Technology, E: Economics, E: Environment, P: Politics) framework. This framework has been utilized in researches predicting the future society for the purpose of considering various aspects. Table 5 provides an overview of scenario concepts with abbreviation and corresponding sector in STEEP framework.

4.2.3. Identifying causal relationships and weights

To identify causal relationships and weights, the FARM is implemented using the Fuzzy Apriori-T software (Coenen, 2008). The impact values of concepts in documents derived matrix *V* from LSA are used as the input of FARM. It can be interpreted as the degree of membership to be particular concept in each document. Table 6 illustrates part of the transaction matrix for FARM. For instance, the degree of membership to be concept JC in document 1 is 0.108, and AE in document 3 is 0.014.

Then, to apply FARM, value of minimum support and minimum confidence are established: $minsupp = 0.4$, $minconf = 0.23$. At first, the 72 association rules are derived using support-confidence framework. The confidence and lift values of them are included in Table 7. When we assess those 72 rules with PA test and MH statistics, 27 rules are not significant and only 45 rules are significant causal rules. As shown in Table 7, the rules marked with $*$ are tested as non-causal relationships and thus excluded.

4.2.4. Finalizing and visualizing scenario model

As was pointed out in the detailed process of this paper, sign of the relationship between concepts constituting the FCM is decided by lift value: sign is positive if lift of rule is higher than 1 and vice versa. Therefore, the sign was given for the 45 rules as shown in Table 8. The table is filled with confidence value with given sign, which means the strength of impact. For instance of the rule EE → AP, the lift value is 0.87 which is lower than 1 and the confidence value is 0.51; thus, causal weight is −0.51. Meanwhile, rules that do not exceed *minsupp* or *minconf* and satisfy PA test are left in space because those two concepts are regarded as not having a meaningful causal relationship.

In the following step, the FCM is constructed using the FCMapper (http://www.fcmapper.net) and Pajek software packages. First, characteristics of vertices and arcs are calculated based on the adjacency matrix, which are generated in Table 8. Second, they can be transformed as a net-file, which Pajek software packages can

Table 4
Defining scenario concepts from semantic textual patterns.

Semantic textual pattern (v_k)	Allocated term (t_i) (impact of term to pattern (u_{ik}))	Definition of scenario concept (C_k)
v_1	Consumer (0.116), customer (0.108), tourism (0.223), growth (0.137), economy (0.15)	Application to tourism
v_2	Automation (0.221), sensor (0.256), network connection (0.098), software (0.175), comfort (0.163), assistant (0.083), internet (0.097)	Usability
v_3	Company (0.131), startup (0.216), university (0.097), laboratory (0.297), investment (0.073), partnership (0.066), entrepreneur (0.093), grid (0.084)	Industry-university cooperation
v_4	Renewable energy (0.137), diesel (0.152), biofuel (0.102), biomass (0.086), geothermal (0.077), petroleum (0.059), gasoline (0.128), hybrid (0.094), photovoltaic (0.065), solar energy (0.058)	Alternative energy technology
v_5	Regulation (0.162), incentive (0.081), policy (0.148), government (0.148), limitation (0.124), standard (0.093), tax reduction (0.117), policy (0.137)	Government regulation
v_6	Engine (0.164), inverter (0.155), magnet (0.076), DC (0.203), AC (0.106), torque (0.059), capacity (0.108), motor (0.122)	Motor technology
v_7	Wireless power (0.182), charger (0.266), recharge (0.243), power transmission (0.097), charger (0.177)	Charging technology
v_8	Transportation (0.19), electric bus (0.15), driver (0.11), passenger (0.106)	Application to public transportation
v_9	Safety (0.344), driverless (0.324), collision (0.12), vibration (0.104), pressure (0.179), security (0.143), stability (0.071), obstacle warning (0.068), monitoring (0.131)	Safety
v_{10}	Economy (0.32), growth (0.15), sales (0.09), investment (0.09), revenue (0.085), GDP (0.181), trade (0.097), import (0.068), export (0.109)	Economic revenue
v_{11}	Energy efficiency (0.138), energy consumption (0.097), efficiency improvement (0.088), energy density (0.067), mileage (0.104)	Energy efficiency
v_{12}	Temperature (0.137), environment (0.185), pollution (0.216), atmosphere (0.207), carbon dioxide emission (0.516), greenhouse gas (0.107), CO_2 (0.114), eco (0.068)	Air pollution
v_{13}	Job (0.311), worker (0.158), manufacturing (0.103), services (0.112), employment (0.217)	Job creation
v_{14}	Lithium battery (0.275), ion battery (0.31), acid battery (0.12), storage (0.124), battery life (0.227), lightweight (0.098), BMS (0.103), lithium ion battery (0.593)	Battery technology
v_{15}	Cost reduction (0.208), incentive (0.094), support (0.103), maintenance cost (0.098)	Costs reduction

Table 5
Scenario concepts classified as STEEP framework.

Type of STEEP framework	Scenario concept (C_k)	Abbreviation
Social (S)	Application to tourism	AT
	Job creation	JC
	Application to public transportation	PT
Technology (T)	Alternative energy technology	AE
	Battery technology	BT
	Motor technology	MT
	Charging technology	CT
	Usability	US
	Safety	SF
Economics (Ec)	Economic revenue	ER
	Costs reduction	CR
Environment (En)	Air pollution	AP
	Energy efficiency	EE
Politics (P)	Industry-university cooperation	IU
	Government regulation	GR

Table 6
Example of transaction data for FARM.

	S			T						Ec		En		P	
	AT	JC	PT	AE	BT	MT	CT	US	SF	ER	CR	AP	EE	IU	GR
Doc 1	0	0.108	0	0.033	0.222	0	0	0	0.175	0	0	0	0	0.048	0.201
Doc 2	0.186	0	0	0	0.664	0	0.177	0	0.13	0	0.059	0	0.035	0.151	0
Doc 3	0	0	0	0.014	0	0.068	0	0.093	0	0	0	0	0	0.156	0
Doc 4	0	0.225	0.142	0	0	0.141	0.102	0	0	0	0	0	0	0.027	0.049
Doc 5	0	0	0	0.009	0.148	0	0	0	0.201	0	0	0	0.061	0.054	0
Doc 6	0	0	0	0	0	0	0.084	0	0	0	0	0.017	0.083	0.192	0
Doc 7	0	0.131	0	0.062	0	0.062	0.102	0.259	0	0.095	0.042	0	0	0	0
Doc 8	0	0	0	0.106	0.104	0.021	0	0	0	0	0	0	0	0	0.058
Doc 9	0	0.174	0.015	0.05	0.21	0	0.206	0	0	0	0	0	0	0	0
Doc 10	0	0	0.07	0.014	0	0	0	0	0	0	0	0	0	0.011	0.106
…															
Doc 2017															

handle with. Lastly, FCM is visualized in Fig. 5 over set of additional options such as size and color. The size of each concept node is determined by the centrality value; and the color and shape of each concept is classified according to the type of STEEP framework explained in the previous pages: red box (society), blue circle (technology), green diamond (economy), yellow circle (environment), and purple triangle (policy). For edges, the solid line means the positive cause and effect relation, otherwise, the dotted line means the negative cause and effect relation.

4.2.5. Analyzing scenarios

4.2.5.1. Static analysis. In the purpose of identifying important scenario concepts, the static analysis gauged the outdegree, indegree, and total centrality, as Table 9. High centrality value means that the concept is important in network or system since concepts not only receive an impact from the concept, but also affect the concept in the FCM. The interesting results are in two concepts: 'EE' and 'AP' have relatively large difference value between indegree and outdegree. 'Air pollution (AP)' can be interpreted as to receive more influenced by other concepts. On the other hand, in the case of 'energy efficiency (EE)', the degree of impact on other concepts is relatively higher than others. Consequently, the important concepts with high centrality value that should be emphasized in scenario planning are derived as: application to tourism (AT), battery technology (BT), air pollution (AP), energy efficiency (EE), and government regulation (GR), as indicated by shaded gray rows in Table 9.

4.2.5.2. Dynamic analysis. The input vectors of dynamic analysis are generated over a combination of binary variation (0 or 1) of five important scenario concepts. As a result derived from the static analysis, the five concepts and corresponding meaning of variations are suggested in the morphology matrix of Table 10. Total of 32 (2^5) combinations of input vectors are possible; however, there are inconsistent combinations of concepts which are excluded. For instance, it cannot be expected to increase energy efficiency without improvements in battery technology. Thus, all the input vectors having a relation of 'BT' and 'EE' in (1, 0) or (0, 1) are excluded.

As a result, we constructed the input vectors of six raw scenarios including basic scenario, which have input vector as (1, 1, 1, 1, 1). Raw scenarios are represented in sequence of each concept's value of variation as below:

- Basic scenario (AP1-PT1-BT1-EE1-GR1)
- Scenario 1: No increase in applying EV to tourism (AP1-PT1-BT1-EE1-GR1)
- Scenario 2: Failure to develop battery technology and energy efficiency (AP1-PT1-BT0-EE0-GR1)
- Scenario 3: Increase of air pollution because of failure of EV (AP1-PT0-BT0-EE0-GR1)
- Scenario 4: Relax of government regulations for vitalizing EV (AP1-PT1-BT1-EE1-GR0)

The output vectors of scenarios are drawn by iterations of multiplying input vector and adjacency matrix until stabilization. The numbers of iteration until stabilization and the input and output values of concepts in each raw scenario are represented in Table 11.

From Table 11, various what-if experiments can imply the impact of scenario concepts on the whole model of future of EV. Comparing the gap between the basic scenario and other scenarios can identify the impact of raw scenarios. In the scenario 1, if the application of EV in the tourism industry decreases, the successive

Table 7
Output association rules and PA test results.

Antecedent	Consequent	Confidence	Lift	Antecedent	Consequent	Confidence	Lift
EE	AP	0.51	0.87	AT	ER	0.29	1.25
CT*	AP*	0.46	0.94	PT*	ER*	0.29	1.01
AT*	AP*	0.44	0.95	EE	GR	0.29	0.71
AP	GR	0.44	1.14	EE	AT	0.28	1.07
EE*	BT*	0.43	1.03	EE	ER	0.28	1.11
PT*	AP*	0.42	0.97	EE	AE	0.28	1.18
GR	AP	0.41	1.14	GR*	AT*	0.27	0.95
CT	PT	0.39	1.15	AT	JC	0.27	1.42
CT	AT	0.38	1.15	SF	PT	0.27	1.02
IU	JC	0.36	1.08	GR	PT	0.27	1.4
BT*	PT*	0.36	1.03	EE	MT	0.27	1.48
PT*	SF*	0.36	1.02	CR	ER	0.27	1.12
US	SF	0.36	1.16	BT	AP	0.27	0.94
AE	AP	0.36	0.89	CR*	AP*	0.27	0.98
BT	EE	0.36	1.03	US*	AT*	0.26	1.04
PT*	AT*	0.35	1.02	CR*	AT*	0.26	1.04
AP	BT	0.34	0.94	IU	BT	0.26	1.01
SF	US	0.34	1.16	BT*	CT*	0.26	1.03
GR	CR	0.34	0.92	JC	ER	0.26	1.33
AP*	AT*	0.33	0.95	AP	ER	0.26	0.87
EE	PT	0.33	1.06	MT*	CR*	0.26	1.03
AT	BT	0.33	1.08	EE	CR	0.26	1.07
PT	CT	0.33	1.15	IU	AP	0.26	0.87
ER	AP	0.33	0.87	AP	AE	0.26	0.89
PT*	IU*	0.33	0.97	AE	BT	0.25	1.05
BT	AT	0.32	1.08	CT*	BT*	0.25	1.03
GR	US	0.32	0.92	GR*	BT*	0.25	1.05
CT	CR	0.32	1.06	BT*	MT*	0.25	1.01
GR*	IU*	0.32	0.96	AE*	CR*	0.25	1.03
PT	US	0.31	1.08	BT*	CR*	0.25	0.98
AT	SF	0.31	0.94	ER*	IU*	0.25	1.03
IU	CR	0.31	1.08	AT	CT	0.24	1.15
AE	EE	0.31	1.18	IU*	CT*	0.24	1.02
AT*	PT*	0.3	1.02	MT	EE	0.24	1.48
GR	SF	0.3	1.11	GR*	JC*	0.23	0.95
PT*	JC*	0.29	0.96	GR	EE	0.23	0.71

* Insignificant rules, given significance level $\alpha = 0.05$, $\chi^2_\alpha = 3.84$.

Table 8
Adjacency matrix.

	S			T						Ec		En		P	
	AT	JC	PT	AE	BT	MT	CT	US	SF	ER	CR	AP	EE	IU	GR
AT		0.27			0.33		0.24		−0.31	0.29					
JC										0.26					
PT							0.33	0.31							
AE					0.25							−0.36	0.31		
BT	0.32											−0.27	0.36		
MT													0.24		
CT	0.38										0.32				
US									0.36						
SF			0.27					0.34							
ER												−0.33			
CR										0.27					
AP				−0.26	−0.34					−0.26					0.44
EE	0.28		0.33	0.28		0.27				0.28	0.26	−0.51			−0.29
IU		0.36			0.26						0.31	−0.26			
GR			0.27					−0.32	0.30		−0.34	0.41	−0.23		

negative impacts arise in the level of job creation (−0.05), battery technology (−0.05), charging technology (−0.04), safety technology (−0.05), economic revenue (−0.05), air pollution (−0.01), and energy efficiency (−0.01). Reversely, we can identify some implications: if electric vehicles are utilized for tourism, related jobs may be available; the application to tourism can cause economic benefits by reducing the level of air pollution due to reduced environmental restoration costs; and since increase of application to tourism promote technological factors, those technologies should not be developed without such motivation.

Scenario 2 shows the negative impacts of failure to increasing energy efficiency caused by the un-development of battery technology: job creation (−0.01), application to public transportation (−0.01), application to tourism (−0.05), economic revenue (−0.01) and air pollution (0.09); that is, the battery technology associated with energy efficiency is key to EV success and market vitality

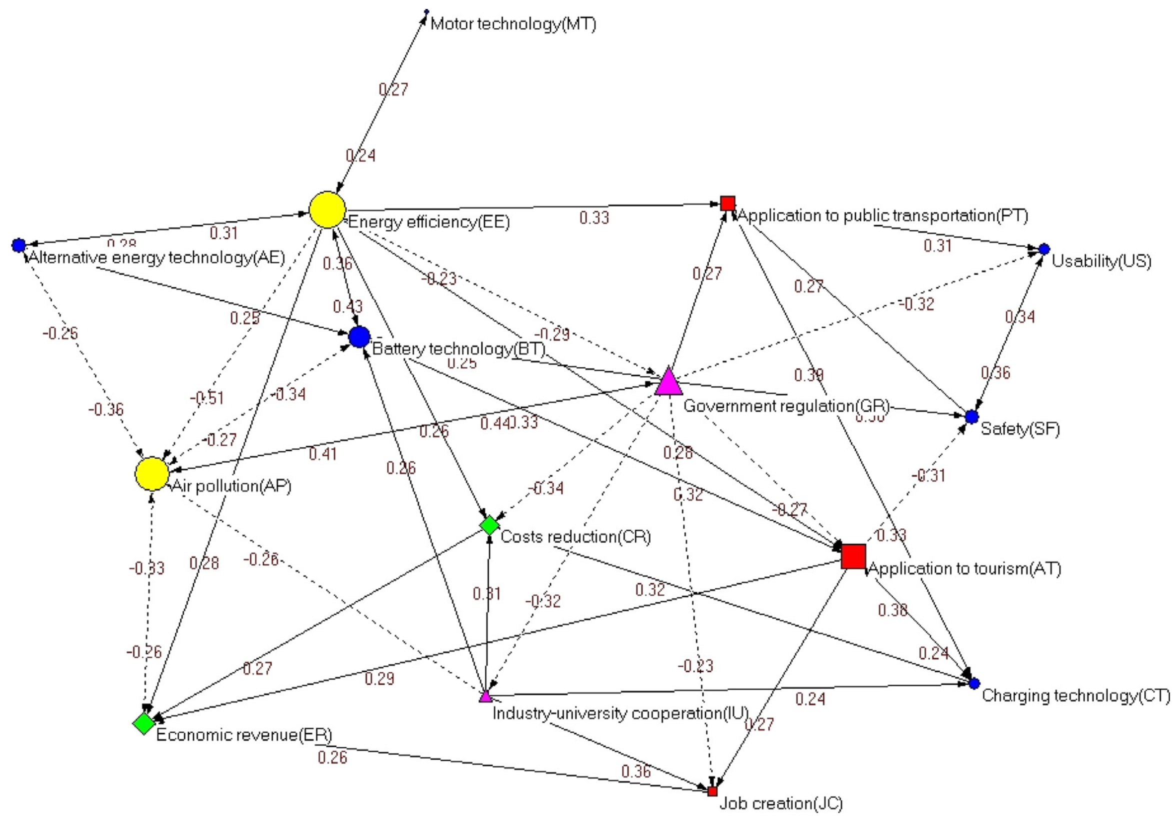


Fig. 5. FCM for EV represented by Pajek: the size of nodes corresponds to the centrality in network, the color and shape of nodes means STEEP framework, and solid/dotted line means positive/negative relationships. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 9
FCM indices and five important concepts marked as gray rows.

STEEP	Concepts	Outdegree	Indegree	Centrality
Social (S)	Application to tourism (AT)	2.18	2.45	4.63
	Job creation (JC)	0.59	1.15	1.74
	Application to public transportation (PT)	2.35	1.92	4.27
Technology (T)	Alternative energy technology (AE)	1.17	0.54	1.71
	Battery technology (BT)	2.07	2.11	4.18
	Motor technology (MT)	0.50	0.52	1.02
	Charging technology (CT)	1.80	1.07	2.87
	Usability (US)	0.62	0.97	1.59
	Safety (SF)	0.61	1.33	1.94
Economics (Ec)	Economic revenue (ER)	0.58	1.65	2.23
	Costs reduction (CR)	0.80	1.99	2.79
Environment (En)	Air pollution (AP)	1.63	3.73	5.36
	Energy efficiency (EE)	2.93	1.14	4.07
Politics (P)	Industry-university cooperation (IU)	1.43	0.90	2.33
	Government regulation (GR)	2.94	0.73	3.67

because it contributes the industrial applications (AT and PT), as well as economic revenue. Although the previous scenarios represent the impact of sole factors, the scenario 3 describes the integrative results of all important scenario concepts. If the all important factors fail to implement EV, expected results are significant as the increase in air pollution probability (0.13), decrease in economic revenue (−0.10) and decrease in associated technologies such as alternative energy technology (−0.05), motor technology (−0.04), charging technology (−0.04) and safety (−0.04), etc. Finally, in the scenario 4, the impact of relax in government regulation such as tax reduction and incentive policy is identified: increase in safety technology (0.05), cost reduction (0.04), and energy efficiency (0.02).

5. Discussion and conclusions

5.1. Practical implications

This paper suggested the method of futuristic data-driven scenario building by incorporating TM and FARM into FCM. The paper insists that futuristic data containing the opinions and discussions on the shape of future from large-participation, as a significant and proper source of a scenario development. In order to apply futur-

Table 10
Morphology matrix.

	AP	PT	BT	EE	GR
Variation A (1)	Increase (1)	Increase (1)	Develop (1)	Increase (1)	Increase (1)
Variation B (0)	No increase (0)	No increase (0)	Not develop (0)	No increase (0)	No increase (0)

Table 11
Dynamic scenarios – input (I) and output (O) vectors.

	Concept	Basic		Scenario 1		Scenario 2		Scenario 3		Scenario 4	
		I	O	I	O	I	O	I	O	I	O
S	Application to tourism (AT)	1	0.64	0	0	1	0.59	0	0	1	0.64
	Job creation (JC)	1	0.59	1	0.54	1	0.58	1	0.54	1	0.59
	Application to public transportation (PT)	1	0.62	1	0.62	1	0.61	1	0.58	1	0.58
T	Alternative energy technology (AE)	1	0.52	1	0.52	1	0.51	1	0.47	1	0.53
	Battery technology (BT)	1	0.59	1	0.54	0	0	0	0	1	0.59
	Motor technology (MT)	1	0.54	1	0.54	1	0.54	1	0.50	1	0.54
	Charging technology (CT)	1	0.59	1	0.55	1	0.59	1	0.55	1	0.59
	Usability (US)	1	0.55	1	0.56	1	0.55	1	0.55	1	0.59
	Safety (SF)	1	0.54	1	0.59	1	0.54	1	0.50	1	0.59
Ec	Economic revenue (ER)	1	0.64	1	0.59	1	0.63	1	0.54	1	0.65
	Costs reduction (CR)	1	0.58	1	0.58	1	0.58	1	0.54	1	0.62
En	Air pollution (AP)	1	0.31	1	0.32	1	0.36	1	0.44	1	0.27
	Energy efficiency (EE)	1	0.60	1	0.59	0	0.54	0	0	1	0.62
P	Industry-university cooperation (IU)	1	0.50	1	0.50	1	0.50	1	0.50	1	0.50
	Government regulation (GR)	1	0.49	1	0.49	1	0.50	1	0.55	0	0
Number of iterations		25		24		22		1		22	

istic data into scenario building, LSA of TM is applied to extract scenario concepts that are defined as the coherent cluster set of keywords; FARM is utilized to identify the causal rules among concepts and measure the weight values of their rules; and finally FCM is constructed and analyzed in terms of static and dynamic viewpoints.

The study is motivated from the prior inductive and deductive developments of FCM-based scenario: the dependence on limited domain knowledge and subjectivity (deductive modeling), and the lack of deriving concept nodes and tendency to utilize historical data (inductive modeling). Considering these points, the proposed method has several advantages. First, the leverage of futuristic data can capture a priori, future-oriented information for scenario, not posteriori information. In our case, in the future of EV, the concepts such as application of tourism and public transportation are future-oriented factors, as if they are extracted from the scenario planning. Second, in terms of variety, the knowledge span can be various as the scenario concepts are widely spanning on STEEP fields. Since large participants from different domains, whether they are experts or ordinary people, freely share any types of future-oriented information, the extraction of cognitive models of these can be said to be based on collective intelligence. Furthermore, the dynamic analysis implements various what-if scenarios, which can help to generate a variety of scenarios. Third, not only weights, but concept nodes of FCMs can be identified from futuristic data. As we involved the LSA to extract scenario concept nodes and the FARM to identify the weights of causal relationship, these algorithms can aid to deal with vast amount of futuristic database and improve the effectiveness and efficiency of scanning knowledge for FCM-based scenario development.

5.2. Limitations of the study

Despite the contributions, there are limitations and rooms to be elaborated in future research. First, the framework considered support, confidence, and lift of FARM as the measure of interestingness of rules; however, other alternative measures can be incorporated such as leverage, conviction, etc. Although we assume that the confidence is the strength of causal dependency, identifying true strength would greatly enhances the value of scenarios. According to Mazlack (2004), confidence is larger than or equal to causal dependence. Thus, there is need for modifying confidence to estimate the strength of causal relationships. Second, although futuristic data can involve multiple stakeholders, we did not separate and integrate models in terms of different viewpoints. Since the FCM has advantage in combining different FCMs, future research can attempt to compare and integrate multiple stakeholders' insights by deriving FCM model based on consensus. Third, the scenario concepts are defined at the level of semantic textual patterns; however, this can be elaborated in more hierarchical ways. For example, the keywords in each semantic textual pattern can be defined as micro-scenario concepts for building more specific and detailed micro-level scenarios. In this case, FCMs are modeled in macro-level FCMs consisting of semantic textual pattern, as well as in micro-level FCMs consisting of keywords. Lastly, the static and dynamic analysis can be further elaborated with advanced techniques. Recently there are emerging studies on improving FCM performances in terms of more flexible static and dynamic analysis. For example, Dias, Hadjileontiadou, Hadjileontiadis, and Diniz (2015) considered influential concepts' contribution to self-sustained cycles in static analysis and time (in) dependence in dynamic analysis. Likewise, the static and dynamic analysis for scenarios can be extended to define centrality for specific context or involve various factors such as and update time and evolving behaviors.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MOE) (No. 2014R1A1A2054064).

References

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on management of data, SIGMOD'93: Vol. 22* (pp. 207–216). New York, New York, USA: ACM Press. http://doi.org/10.1145/170035.170072 .

Amer, M., & Daim, T. (2013). Expert judgment quantification. In T. Daim, T. Oliver, & J. Kim (Eds.), *Research and technology management in the electricity industry* (pp. 31–65). London: Springer.
Amer, M., Daim, T. U., & Jetter, A. (2013). A review of scenario planning. *Futures, 46*, 23–40 http://doi.org/10.1016/j.futures.2012.10.003.
Amer, M., Daim, T. U., & Jetter, A. (2013). Application of fuzzy cognitive map for the development of scenarios: A case study of Wind Energy Deployment. *Research and technology management in the electricity industry* (pp. 129–159). London: Springer.
Berry, M. W., & Kogan, J. (2010). *Text mining: applications and theory*. John Wiley & Sons.
Biloslavo, R., & Dolinšek, S. (2010). Scenario planning for climate strategies development by integrating group delphi, AHP and dynamic fuzzy cognitive maps. *Foresight, 12*(2), 38–48 http://doi.org/10.1108/14636681011035771.
Bishop, P., Hines, A., & Collins, T. (2007). The current state of scenario development: An overview of techniques. *Foresight, 9*(1), 5–25 http://doi.org/10.1108/14636680710727516.
Brin, S., Motwani, R., Ullman, J. D., & Tsur, S. (1997). Dynamic itemset counting and implication rules for market basket data. *ACM SIGMOD Record, 26*(2), 255–264 http://doi.org/10.1145/253262.253325.
Cachia, R., Compañó, R., & Costa, O. Da. (2007). Grasping the potential of online social networks for foresight. *Technological Forecasting and Social Change, 74*(8), 1179–1203.
Chen, Y., Mazlack, L. J., Minai, A. A., & Lu, L. J. (2015). Inferring causal networks using fuzzy cognitive maps and evolutionary algorithms with application to gene regulatory network reconstruction. *Applied Soft Computing, 37*, 667–679 http://doi.org/10.1016/j.asoc.2015.08.039.
Retrieved from:Accessed 16.11.15 Coenen, F.. *The LUCS-KDD fuzzy Apriori-T software*. Retrieved from:Accessed 16.11.15 https://cgi.csc.liv.ac.uk/~frans/KDD/Software/FuzzyAprioriT/fuzzyAprioriT.html .
Dias, S. B., Hadjileontiadou, S. J., Hadjileontiadis, L. J., & Diniz, J. A. (2015). Fuzzy cognitive mapping of LMS users' quality of interaction within higher education blended-learning environment. *Expert Systems with Applications, 42*(21), 7399–7423 http://doi.org/10.1016/j.eswa.2015.05.048.
Drew, S. (2006). Building technology foresight: Using scenarios to embrace innovation. *European Journal of Innovation Management, 9*(3), 241–257.
Dumais, S. T. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology, 38*(1), 188–230 http://doi.org/10.1002/aris.1440380105.
Farzanyar, Z., & Kangavari, M. (2012). Efficient mining of fuzzy association rules from the pre-processed dataset. *Computing and Informatics, 31*(2), 331–347.
Ferreira, F. A. F., Jalali, M. S., & Ferreira, J. J. M. (2015). Integrating qualitative comparative analysis (QCA) and fuzzy cognitive maps (FCM) to enhance the selection of independent variables. *Journal of Business Research, 69*(4), 1471–1478 http://doi.org/10.1016/j.jbusres.2015.10.127.
Hirsch, S., Burggraf, P., & Daheim, C. (2013). Scenario planning with integrated quantification – Managing uncertainty in corporate strategy building. *Foresight, 15*(5), 363–374 http://doi.org/10.1108/FS-09-2012-0064.
Jetter, A. J., & Kok, K. (2014). Fuzzy cognitive maps for futures studies—A methodological assessment of concepts and methods. *Futures, 61*, 45–57 http://doi.org/10.1016/j.futures.2014.05.002.
Jetter, A., & Schweinfort, W. (2011). Building scenarios with fuzzy cognitive maps: An exploratory study of solar energy. *Futures, 43*(1), 52–66 http://doi.org/10.1016/j.futures.2010.05.002.
Jin, Z., Li, J., Liu, L., Le, T. D., Sun, B., & Wang, R. (2012). Discovery of causal rules using partial association. In *Proceedings of the 2012 IEEE 12th international conference on data mining* (pp. 309–318). IEEE. http://doi.org/10.1109/ICDM.2012.36.
Kahn, H., & Wiener, A. J. (1967). *The year 2000: A framework for speculation on the next thirty-three years*. New York: Macmillan.
Keller, J., & von der Gracht, H. A. (2014). The influence of information and communication technology (ICT) on future foresight processes – Results from a delphi survey. *Technological Forecasting and Social Change, 85*, 81–92 http://doi.org/10.1016/j.techfore.2013.07.010.
Khan, M. S., & Quaddus, M. (2004). Group decision support using fuzzy cognitive maps for causal reasoning. *Group Decision and Negotiation, 13*(5), 463–480 http://doi.org/10.1023/B:GRUP.0000045748.89201.f3.
Kim, J., & Park, Y. (2014). Supporting technology foresight for disruptive innovation : Keyword-based visual analysis for futuristic data. In *Proceedings of the 2014 Asia Pacific industrial engineering & management systems conference*.
Kok, K. (2009). The potential of fuzzy cognitive maps for semi-quantitative scenario development, with an example from Brazil. *Global Environmental Change, 19*(1), 122–133 http://doi.org/10.1016/j.gloenvcha.2008.08.003.
Kosko, B. (1986). Fuzzy cognitive maps. *International Journal of Man-Machine Studies, 24*(1), 65–75 http://doi.org/10.1016/S0020-7373(86)80040-2.
Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*(2-3), 259–284 http://doi.org/10.1080/01638539809545028.
Lee, D. H., & Lee, H. (2015). Construction of holistic fuzzy cognitive maps using ontology matching method. *Expert Systems with Applications, 42*(14), 5954–5962 http://doi.org/10.1016/j.eswa.2015.03.020.
Li, J., Liu, L., & Le, T. D. (2015). Causal rule discovery with partial association test. *Practical approaches to causal relationship exploration* (pp. 33–50). Cham: Springer International Publishing http://doi.org/10.1007/978-3-319-14433-7.
Lin, C. W., Chen, S. H., & Tzeng, G. H. (2009). Constructing a cognition map of alternative fuel vehicles using the DEMATEL method. *Journal of Multi-Criteria Decision Analysis, 16*(1-2), 5–19 http://doi.org/10.1002/mcda.438.
Lin, F. R., Hsieh, L. S., & Chuang, F. T. (2009). Discovering genres of online discussion threads via text mining. *Computers & Education, 52*(2), 481–495 http://doi.org/10.1016/j.compedu.2008.10.005.
Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*(4), 719–748.
Markmann, C., von der Gracht, H. A., Keller, J., & Kroehl, R. (2012). Collaborative foresight as a means to face future risks – An innovative platform conception. In *Proceedings of the 9th international ISCRAM conference* (pp. 1–5).
Mazlack, L. J. (2004). Using association rules without understanding their underlying causality reduces their decision value. In *Proceedings of the 2004 international conference on machine learning and applications, 2004* (pp. 312–319). IEEE. http://doi.org/10.1109/ICMLA.2004.1383529.
Mietzner, D., & Reger, G. (2005). Advantages and disadvantages of scenario approaches for strategic foresight. *International Journal of Technology Intelligence and Planning, 1*(2), 220–239.
Motlagh, O., Jamaludin, Z., Tang, S. H., & Khaksar, W. (2015). An agile FCM for real-time modeling of dynamic and real-life systems. *Evolving Systems, 6*(3), 153–165 http://doi.org/10.1007/s12530-013-9077-6.
Pang, A. S. K. (2010). Futures 2.0: Rethinking the discipline. *Foresight, 12*(1), 5–20 http://doi.org/10.1108/14636681011020191.
Papageorgiou, E. I. (2012). Learning algorithms for fuzzy cognitive maps—a review study. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 42*(2), 150–163 http://doi.org/10.1109/TSMCC.2011.2138694.
Papageorgiou, E. I., & Salmeron, J. L. (2013). A review of fuzzy cognitive maps research during the last decade. *IEEE Transactions on Fuzzy Systems, 21*(1), 66–79 http://doi.org/10.1109/TFUZZ.2012.2201727.
Raford, N. (2015). Online foresight platforms: Evidence for their impact on scenario planning & strategic foresight. *Technological Forecasting and Social Change, 97*, 65–76 http://doi.org/10.1016/j.techfore.2014.03.008.
Ritchey, T. (2006). Problem structuring using computer-aided morphological analysis. *Journal of the Operational Research Society, 57*(7), 792–801.
Salmeron, J. L., Vidal, R., & Mena, A. (2012). Ranking fuzzy cognitive map based scenarios with TOPSIS. *Expert Systems with Applications, 39*(3), 2443–2450 http://doi.org/10.1016/j.eswa.2011.08.094.
Schatzmann, J., Schäfer, R., & Eichelbaum, F. (2013). Foresight 2.0 – Definition, overview & evaluation. *European Journal of Futures Research, 1*(1), 1–15 http://doi.org/10.1007/s40309-013-0015-4.
Schoemaker, P. J. H. (1995). Scenario planning: A tool for strategic thinking. *Sloan Management Review, 36*(2), 25–40.
Soler, L. S., Kok, K., Camara, G., & Veldkamp, A. (2012). Using fuzzy cognitive maps to describe current system dynamics and develop land cover scenarios: A case study in the Brazilian Amazon. *Journal of Land Use Science, 7*(2), 149–175 http://doi.org/10.1080/1747423X.2010.542495.
Srikant, R., & Agrawal, R. (1996). Mining quantitative association rules in large relational tables. *ACM SIGMOD Record, 25*(2), 1–12.
Stach, W., Kurgan, L., & Pedrycz, W. (2010). Fuzzy cognitive maps. In M. Glykas (Ed.), *Studies in fuzziness and soft computing: Vol. 247* (pp. 23–41). Berlin, Heidelberg: Springer http://doi.org/10.1007/978-3-642-03220-2.
Stach, W., Kurgan, L., Pedrycz, W., & Reformat, M. (2005). Genetic learning of fuzzy cognitive maps. *Fuzzy Sets and Systems, 153*(3), 371–401 http://doi.org/10.1016/j.fss.2005.01.009.
Thorleuchter, D., Scheja, T., & Van den Poel, D. (2014). Semantic weak signal tracing. *Expert Systems with Applications, 41*(11), 5009–5016 http://doi.org/10.1016/j.eswa.2014.02.046.
Thorleuchter, D., & Van den Poel, D. (2013). Technology classification with latent semantic indexing. *Expert Systems with Applications, 40*(5), 1786–1795 http://doi.org/10.1016/j.eswa.2012.09.023.
Thorleuchter, D., & Van den Poel, D. (2014). Semantic compared cross impact analysis. *Expert Systems with Applications, 41*(7), 3477–3483 http://doi.org/10.1016/j.eswa.2013.10.051.
Tsadiras, A., & Bassiliades, N. (2013). RuleML representation and simulation of fuzzy cognitive maps. *Expert Systems with Applications, 40*(5), 1413–1426 http://doi.org/10.1016/j.eswa.2012.08.035.
Tsadiras, A. K. (2008). Comparing the inference capabilities of binary, trivalent and sigmoid fuzzy cognitive maps. *Information Sciences, 178*(20), 3880–3894 http://doi.org/10.1016/j.ins.2008.05.015.
Yaman, D., & Polat, S. (2009). A fuzzy cognitive map approach for effect-based operations: An illustrative case. *Information Sciences, 179*(4), 382–403 http://doi.org/10.1016/j.ins.2008.10.013.
Yongmei, G., & Fuguang, B. (2015). The research on measure method of association rules mining. *International Journal of Database Theory and Application, 8*(2), 245–258.
Yoon, B., Phaal, R., & Probert, D. (2008). Morphology analysis for technology roadmapping: Application of text mining. *R&D Management, 38*(1), 51–68 http://doi.org/10.1111/j.1467-9310.2007.00493.x.
Zhang, Y., Guo, Y., Wang, X., Zhu, D., & Porter, A. L. (2013). A hybrid visualisation model for technology roadmapping: Bibliometrics, qualitative methodology and empirical study. *Technology Analysis & Strategic Management, 25*(6), 707–724 http://doi.org/10.1080/09537325.2013.803064.
Zhu, D., & Porter, A. (2002). Automated extraction and visualization of information for technological intelligence and forecasting. *Technological Forecasting and Social Change, 69*(5), 495–506.