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.