Expert Systems with Applications 40 (2013) 429–438

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Idea discovery: A scenario-based systematic approach for decision making in market innovation

Hao Wang *, Yukio Ohsawa

Department of Systems Innovation, The University of Tokyo, Tokyo 113-8656, Japan

ARTICLE INFO

Keywords:
Galaxy
Idea discovery
IdeaGraph
Dynamic discovery

ABSTRACT

A new trend of researches on knowledge discovery and chance discovery is to identify human insights through data synthesis rather than to discover facts through data analysis. In this paper, we propose a systematic approach named idea discovery which is committed to turning data into effective human insights. Idea discovery focuses on dynamic and sustainable process for high-quality ideas cultivation, construction, integration and evaluation through human–computer and human–human interaction. It mainly relies on latent information and its dynamic changes to drive ideas creation, integration and evaluation during sustainable creativity process. The process of idea discovery is in accordance with a dynamic model which contains two key components: (1) mining algorithms to turn data into scenario maps for eliciting human insights; (2) scenario-based creativity support activities towards actionable ideas generation. An intelligence system called *Galaxy* integrated with *IdeaGraph* algorithm has been developed to support the dynamic process of idea discovery. A case study in an automobile company has validated the effectiveness of proposed method and system.

1. Introduction

Text-based knowledge discovery is increasingly playing an important role in scientific discovery. It is helpful for researchers to manage valuable information and identify facts in text data that would solve practical problems. On the other hand, early researches on literature-based discovery (LBD) are focused on the scientific field of biomedicine. Swanson (1986) made a new discovery in text literatures and proposed a hypothesis that fish oil could cure Raynaud's disease. Finally, he tested the hypothesis experimentally and clinically (Swanson, 1986, 1987, 1988, 1989). To solve the complexity of text processing, information retrieval (IR) and information extraction (IE) are used to develop system tools to support discovery process (Gordon & Lindsay, 1996; Lindsay & Gordon, 1999; Swanson & Smalheiser, 1997, 1998; Weeber, Klein, de Jong-van den Berg, & Vos, 2001). For text-based understanding, information extraction potentially offers better support than information retrieval as it can extract useful information and relations expressed in the text documents. It is a good way to interpret extracted information by visualizing terms and relationships graphically (Mack & Hehenberger, 2002).

Text-based knowledge discovery is also widely applied in other areas. For instance, criminal investigators may make use of text data to uncover evidence. A general framework is proposed by combining information extraction techniques with visual exploration techniques to provide an approach to make evidence discovery (Chen et al., 2004; Louris & Engelbrecht, 2011). In business, patent-based discovery is used by enterprises or stakeholders to timely be aware of the situation and direction of current technologies so as to rapidly adjust their market strategies (Chen et al., 2004; Fattori, Pedrazzi, & Turra, 2003; Losiewicz, Oard, & Kostoff, 2000).

Chance Discovery, proposed by Ohsawa in 2000, is a relatively new research field as an extension of text-based knowledge discovery. It is a human–computer interaction process to detect rare but important chances for decision making. A chance means to understand an unnoticed event or situation which might be uncertain but significant for a decision (Ohsawa & McBurney, 2003). A core visualization tool called *KeyGraph* can generate scenario map to aid human's value cognition the double-helix process of chance discovery. In fact, *KeyGraph* is a keyword extraction algorithm from a single document using co-occurrence graph (Ohsawa, Benson, & Yachida, 1998). That is, a document is represented as a graph where each node corresponds to a term and each edge means the co-occurrence of two terms. Based on the segmentation of a graph into clusters, *KeyGraph* extracts keywords by selecting the terms which strongly co-occurs with clusters. Fig. 1 shows a scenario map visualized by *KeyGraph*. The red[1] nodes are considered as chance candidates because they act as a bridge linking different sub-scenarios. Human are required to understand and interpret the

* Corresponding author.
E-mail address: wanghao0423@hotmail.com (H. Wang).

[1] For interpretation of colour in Figs. 1 and 7, the reader is referred to the web version of this article.

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.07.044

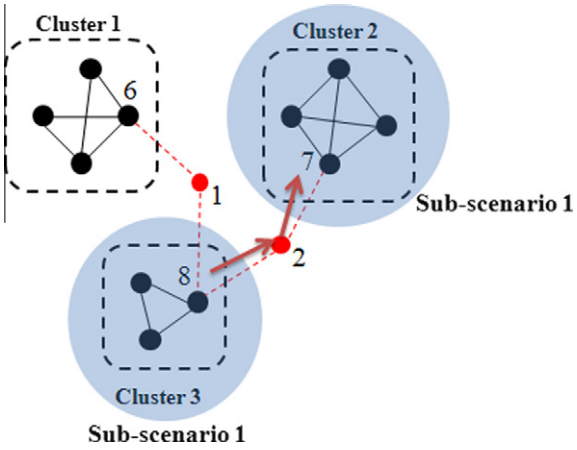

Fig. 1. A scenario map visualized by *KeyGraph*.

importance of chance candidates that can make the situation transfer from one sub-scenario to another. Later, Ohsawa (2005) proposed a breaking-through method named data crystallization where dummy nodes representing invisible events are inserted into the processed text data, and then new data is visualized by *KeyGraph*. However, the complex algorithm and graph obtained were hard for users to understand, thus Maeno and Ohsawa (2007) subsequently present a new method, human–computer interactive annealing, for revealing latent structures and discovering dark events. Based on chance discovery, Hong (2009) proposes an interactive human–computer process model called Qualitative Chance Discovery Model (QCDM) to extract more accurate data representation in scenario map for decision making on potential chances.

In recent years, a scenario map generated by *KeyGraph* with data crystallization has been applied as the game board in Innovators Market Game ® (IMG) and Innovators Marketplace ® (IM) for innovative chance discovery (Ohsawa, 2009; Ohsawa, Okamoto, Takahashi, & Nishihara, 2010, 2012). In particular, human cognition of *KeyGaph* scenario map has been expanded from sensing bridge-like chances to creating chances through combining black nodes, see Fig. 2. Wang and Ohsawa (2012a, 2012b) have proposed a systematic approach with IMG for ideas creation and evaluation, and have applied such an approach in product design. To improve IMG, a customer-centric creativity support techique, 4W-IMG, has been proposed (Wang & Ohsawa, 2011).

Although many relevant researches have successfully achieved knowledge discovery and business chance discovery for decision

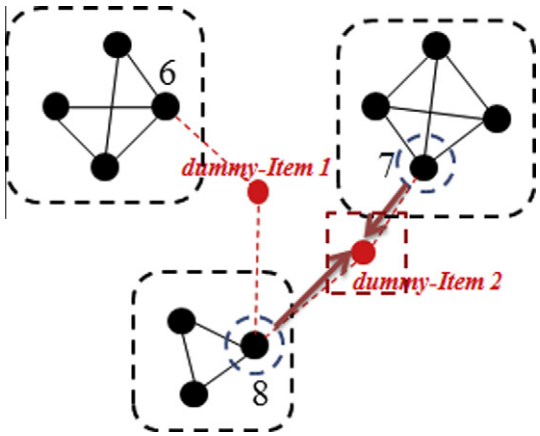

Fig. 2. Creating a chance by combinational thinking in IMG.

making, a new trend in discovery researches is to identify true human insights through data synthesis rather than data analysis.

In this research, we propose a novel systematic approach named idea discovery which extends static discovery into dynamic discovery in three dimensions. A hypothesis is proposed that latent information would be captured across different dimensions. Section 2 summarizes current discovery pattern and introduces the process of dynamic discovery. Idea discovery with a dynamic model is presented in Section 3. In Section 4, a mining algorithm called *IdeaGraph* as a key component of idea discovery is presented for discovering more latent information (event and their relations). A case study in automobile industry is described in Section 5 and the conclusion is in Section 6.

2. Discovery pattern and dynamic discovery

Previous researches on text-based knowledge discovery and chance discovery both combine two complementary processes. One is the process of Computer's text data mining and visualization; the other is the process of human's cognition, understanding, interpretation and innovation. Fig. 3 shows a basic framework of current discovery pattern.

Dynamic discovery is commonly used by enterprises to evaluate and understand technology trends through patent data analysis, and eventually achieve a strategic advantage. Kim, Suh, and Park (2008) proposes a patent map visualization method, but it fails to automatically track the changes of patent trends in different periods. Shih, Liu, and Hsu (2010) overcome the problem and propose a patent trend mining method to automatically identify changes of patent trends without specialist knowledge.

In this paper, dynamic discovery focuses on discovering dynamic changes of event points and their relationship/links through the comparison of two consecutive scenario maps in time series. Fig. 4 shows the dynamic change process of event points and their relationship in different scenario maps, i.e., the status of event points or their relationship changing from nonexistence/high frequency at T0 to low frequency at T1 to high frequency/nonexistence at T2.

3. Idea discovery

Idea discovery, as an extension and evolution of previous discovery researches, is a dynamic and sustainable process for high-quality ideas cultivation, construction, generation and evaluation through human–computer and human–human interaction. Idea discovery not only works on rare and important event points, but also focuses on latent and significant event relationship and the dynamic changes of these events and their relationship. Therefore, idea discovery is committed to digging up latent information (event points and their relationship) and its dynamic changes through static and dynamic discovery, for more actionable ideas creation, integration and evaluation.

Fig. 5 reveals a dynamic model of idea discovery process and the details are presented as below:

Step 1: Data gathering and preprocessing. Determine the objective of task and select relevant data. And then text data is preprocessed into a group of basket data sets denoted by D, each row of which is a basket data set. For example, $P1$, $P2$, $P3$ and $P4$, these four items constitute a basket data set.

$$D = P1, P2, P3, P4$$
$$P2, P7, P5$$
$$P3, P10, P6, P9, P5$$
$$\cdots\cdots\cdots\cdots\cdots$$