Knowledge-Based Systems 88 (2015) 264–272

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A data-driven approximate causal inference model using the evidential reasoning rule

Yue Chen [a], Yu-Wang Chen [b,*], Xiao-Bin Xu [c], Chang-Chun Pan [a], Jian-Bo Yang [b], Gen-Ke Yang [a]

[a] Department of Automation, Shanghai Jiao Tong University, China
[b] Manchester Business School, The University of Manchester, Manchester M15 6PB, UK
[c] Institute of System Science and Control Engineering, School of Automation, Hangzhou Dianzi University, Hangzhou, China

ARTICLE INFO

Article history:
Received 5 January 2015
Received in revised form 17 May 2015
Accepted 21 July 2015
Available online 29 July 2015

Keywords:
Evidential reasoning
Bayesian inference
Belief distribution
Approximate causal inference

ABSTRACT

This paper aims to develop a data-driven approximate causal inference model using the newly-proposed evidential reasoning (ER) rule. The ER rule constitutes a generic conjunctive probabilistic reasoning process and generalises Dempster's rule and Bayesian inference. The belief rule based (BRB) methodology was developed to model complicated nonlinear causal relationships between antecedent attributes and consequents on the basis of the ER algorithm and traditional IF-THEN rule-based systems, and in essence it keeps methodological consistency with Bayesian Network (BN). In this paper, we firstly introduce the ER rule and then analyse its inference patterns with respect to the bounded sum of individual support and the orthogonal sum of collective support from multiple pieces of independent evidence. Furthermore, we propose an approximate causal inference model with the kernel mechanism of data-based approximate causal modelling and optimal learning. The exploratory approximate causal inference model inherits the main strengths of BN, BRB and relevant techniques, and can potentially extend the boundaries of applying approximate causal inference to complex decision and risk analysis, system identification, fault diagnosis, etc. A numerical study on the practical pipeline leak detection problem demonstrates the applicability and capability of the proposed data-driven approximate causal inference model.

1. Introduction

The evidential reasoning (ER) rule has been established recently to combine multiple pieces of independent evidence conjunctively with weights and reliabilities [34]. Through the implementation of the orthogonal sum operation on weighted belief distributions with reliabilities, the ER rule takes into account both individual and collective support from two pieces of evidence in a rational way, and it constitutes a generic conjunctive probabilistic reasoning process or a generalised Bayesian inference process [34,35]. In inheritance of the basic probabilistic properties of being associative and commutative, the ER rule can be easily used to aggregate multiple pieces of evidence recursively. It is expected that the ER rule can further extend the boundaries of existing Bayesian inference methodology and provide a scientific way of reasoning with various probabilistic uncertainties. The ER rule advances the seminal Dempster–Shafer (D–S) theory of evidence [7,16] and the original ER algorithm [29,33]. It has been proved that (1) Dempster's combination rule is a special case of the ER rule when each piece of evidence is fully reliable, and (2) the ER algorithm is also a special case when the reliability of each piece of evidence is assumed to be equal to its normalised weight. In a theoretical sense, the reliability of each piece of evidence is used to measure its inherent quality of the information source, and contrarily the normalised weight reflects its relative importance compared with other pieces of evidence [34]. Previously it was widely accepted that the D–S theory of evidence is one of the most prominent work to generalise Bayesian inference, which consists of a rigorous probabilistic reasoning process [16,34]. In the D–S theory, a frame of discernment is defined by a set of mutually exclusive and collectively exhaustive hypotheses. It is assumed that basic probabilities can be assigned to not only singleton hypotheses but also any subsets of hypotheses. As a result, each piece of evidence is profiled by a belief distribution on the power set of the frame of discernment. Correspondingly, belief distribution is a generalisation of conventional probability distribution in which basic probabilities are only assigned to singleton hypothesis. However, when combining highly or completely conflicting evidence, Dempster's rule combination was found to generate counter-intuitive results [36,27].

* Corresponding author. Tel.: +44 161 2756345.
E-mail address: yu-wang.chen@mbs.ac.uk (Y.-W. Chen).

http://dx.doi.org/10.1016/j.knosys.2015.07.026
0950-7051/

Thereafter, much work has been undertaken to resolve the issue and design new combination rules [18,9,28]. The ER algorithm was originally presented in the context of multiple criteria decision analysis. The holistic approach consists of the belief structure for modelling various types of uncertainty [31,25], the rule and utility based information transformation techniques [29], and the ER algorithm for information aggregation [33], etc. In the past twenty years, the ER algorithm has been widely applied to many system and decision analysis problems as surveyed by Xu [25]. It has also been extended to multi-criteria fuzzy decision-making problems [22,23], fuzzy failure mode and effects analysis [13], rule-base evidential reasoning [8], group decision analysis [11], medical diagnosis [24], and so on. Furthermore, the ER algorithm was introduced to extend traditional If-Then rule based systems to belief rule based (BRB) systems [30]. The BRB methodology employs the informative belief structure to represent various types of information and knowledge with uncertainties and shows superior capability of approximating complicated nonlinear causal relationships across a wide variety of application areas, including fault diagnosis, system identification, risk and decision analysis [26,1,37,6,4].

Given that the ER rule has explicitly generalised the D–S theory of evidence and the original ER algorithm, it becomes perfectly logical and also extremely important to revisit and further improve those techniques which were previously developed from the latter two methods. In this paper, we aim to conduct some exploratory research of building a data-driven approximate causal inference model using the ER Rule and sharpening the edges of the ER and BRB methodologies. The rest of the paper is organised as follows: in Section 2, the inference patterns of the ER rule with respect to the bounded sum of individual support and the orthogonal sum of collective support from multiple pieces of evidence are analysed on the basis of its fundamentals. In Section 3, an approximate causal inference model using the ER rule is explored in view of data-based causal modelling and optimal learning. A numerical study is conducted to illustrate the applicability of the proposed data-driven approximate causal inference model in Section 4. Some concluding remarks are presented in Section 5.

2. The ER rule for inference

2.1. Brief introduction of the ER rule

Suppose a frame of discernment $\Theta = \{\theta_1, \ldots, \theta_N\}$ is a set of mutually exclusive and collectively exhaustive hypotheses, with $\theta_n \cap \theta_m = \emptyset$ for any $n, m \in \{1, \ldots, N\}$ and $n \neq m$ where $\emptyset$ is an empty set. The power set of Θ, denoted by $P(\Theta)$ or 2^Θ, consists of 2^N subsets of Θ as follows

$$P(\Theta) = 2^\Theta = \{\emptyset, \theta_1, \ldots, \theta_N, \{\theta_1, \theta_2\}, \ldots, \{\theta_1, \theta_N\}, \ldots, \{\theta_1, \theta_{N-1}\}, \Theta\} \tag{1}$$

In the ER rule, a piece of evidence e_i is profiled by the following belief distribution.

$$e_i = \left\{(\theta, p_{\theta,i}), \forall \theta \subseteq \Theta, \sum_{\theta \subseteq \Theta} p_{\theta,i} = 1\right\} \tag{2}$$

where $p_{\theta,i}$ denotes the degree of belief to which the evidence e_i supports proposition θ being any element of $P(\Theta)$ except for the empty set. $(\theta, p_{\theta,i})$ is referred to as a focal element of e_i if $p_{\theta,i} > 0$. Specifically, the degree of belief assigned exactly to the complete set Θ reflects the degree of global ignorance, and to a smaller subset of Θ except for any singleton proposition measures the degree of local ignorance. If there is no local or global ignorance, the belief distribution reduces to a classical probability distribution [34].

Each piece of evidence e_i is also associated with a weight and a reliability, denoted by w_i and r_i respectively. It is worth noting that weight and reliability are not differentiated clearly in many information aggregation methods [17,34]. In the ER framework, the weight is used to reflect the relative importance of a piece of evidence in comparison with other evidence, and nevertheless the reliability is the inherent property of the evidence and sets the degree of support for a proposition.

As a result, there are mainly three elements to be taken into account when combining a piece of evidence with other evidence: its belief distribution, weight and reliability. The reasoning process in the ER rule is achieved by defining a weighted belief distribution with reliability [34].

$$m_i = \left\{(\theta, \widetilde{m}_{\theta,i}), \forall \theta \subseteq \Theta; (P(\Theta), \widetilde{m}_{P(\Theta),i})\right\} \tag{3}$$

where $\widetilde{m}_{\theta,i}$ measures the degree of support for θ from e_i with taking into account all the three elements.

$$\widetilde{m}_{\theta,i} = \begin{cases} 0, & \theta = \emptyset \\ c_{rw,i} m_{\theta,i}, & \theta \subseteq \Theta, \theta \neq \emptyset \\ c_{rw,i}(1 - r_i), & \theta = P(\Theta) \end{cases} \quad \text{or} \quad \widetilde{m}_{\theta,i} = \begin{cases} 0, & \theta = \emptyset \\ \widetilde{w}_i p_{\theta,i}, & \theta \subseteq \Theta, \theta \neq \emptyset \\ 1 - \widetilde{w}_i, & \theta = P(\Theta) \end{cases} \tag{4}$$

where $m_{\theta,i} = w_i p_{\theta,i}$ and $c_{rw,i} = 1/(1 + w_i - r_i)$. The normalisation factor $c_{rw,i}$ determines $\sum_{\theta \subseteq \Theta} \widetilde{m}_{\theta,i} + \widetilde{m}_{P(\Theta),i} = 1$. Implicitly, a new hybrid weight $\widetilde{w}_i = c_{rw,i} w_i = w_i/(1 + w_i - r_i)$ is used to calculate $\widetilde{m}_{\theta,i}$ from the original belief degree $p_{\theta,i}$, and $\widetilde{m}_{P(\Theta),i} = 1 - \widetilde{w}_i$. The residual support $\widetilde{m}_{P(\Theta),i} = 0$, when $r_i = 1$.

Given the definition of the weighted belief distribution with reliability, the new ER rule can then be used to combine multiple pieces of evidence recursively. Without loss of generality, the combined degrees of belief to which two pieces of independent evidence e_i and e_j jointly support proposition θ, denoted by $p_{\theta,e(2)}$, can be generated by the orthogonal sum of the weighted belief distributions with reliability (i.e., m_i and m_j) as follows

$$p_{\theta,e(2)} = \begin{cases} 0, & \theta = \emptyset \\ \dfrac{\widehat{m}_{\theta,e(2)}}{\sum_{D \subseteq \Theta} \widehat{m}_{D,e(2)}}, & \theta \subseteq \Theta, \theta \neq \emptyset \end{cases} \tag{5}$$

$$\widehat{m}_{\theta,e(2)} = \left[(1 - r_j) m_{\theta,i} + (1 - r_i) m_{\theta,j}\right] + \sum_{B \cap C = \theta} m_{B,i} m_{C,j}, \quad \forall \theta \subseteq \Theta \tag{6}$$

There are mainly two terms in the equation above. The first square bracket term is regarded as the bounded sum of individual support on proposition θ from each of the two pieces of evidence e_i and e_j. $(1 - r_i)$ reflects the unreliability of evidence e_i, and it sets a bound within which e_j can play a limited role. Here we take two extreme cases as examples. When evidence e_i is fully reliable, i.e., $(1 - r_i) = 0, (1 - r_i) m_{\theta,j} = 0$ and the individual support from evidence e_j will not be counted at all. When evidence e_i is fully unreliable, i.e., $(1 - r_i) = 1, (1 - r_i) m_{\theta,j} = m_{\theta,j}$ and the individual support from evidence e_j will be counted completely. The second term is regarded as the orthogonal sum of collective support from both pieces of evidence e_i and e_j, measuring the degree of all intersected support on proposition θ.

2.2. Inference analysis of the ER rule

As introduced previously, the ER rule generalises a few special cases, which can essentially be characterised by the three elements of evidence. Firstly, the ER rule reduces to Bayesian inference given that each piece of evidence is formulated by a probability distribution, or a so-called belief distribution without local or global ignorance. Secondly, with regard to evidence weight and reliability, there are two possible scenarios: (i) The ER rule turns into the