Applied Soft Computing 11 (2011) 3534–3545

Contents lists available at ScienceDirect

Applied Soft Computing

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

Fuzzy-inferenced decisionmaking under uncertainty and incompleteness

Lily R. Liang[a,*], Carl G. Looney[b], Vinay Mandal[c]

[a] Department of Computer Science & Information Technology, University of the District of Columbia, Washington, DC 20008, USA
[b] Department of Computer Science & Engineering, University of Nevada, Reno, NV 89557, USA
[c] Department of Computer Science, Wayne State University, Detroit, MI 48203, USA

ARTICLE INFO

Article history:
Received 20 September 2009
Received in revised form
17 September 2010
Accepted 17 January 2011
Available online 22 January 2011

Keywords:
Fuzzy-belief-state
Fuzzy-belief-state base
Association rule mining
Fuzzy association rules
Decisionmaking
Uncertainty
Incompleteness

ABSTRACT

An outstanding problem is how to make decisions with uncertain and incomplete data from disparate sources without NP-hard algorithms. Here we introduce a new reasoning methodology, fuzzy-inferenced decisionmaking (*FIND*), to solve this problem in polynomial time. In this methodology, a fuzzy-belief-state base (FBSB) is created from historical data of the states of a system by clustering the set of values for each state variable into three clusters upon whose center fuzzy set membership functions LOW, MEDIUM and HIGH are defined. The FBSB is mined for fuzzy association rules using the fuzzy set memberships to infer values for the missing data via these rules. When given an incomplete and uncertain observation of the system state, the observed state is completed via fuzzy association rules. Then each case in the FBSB is matched against the inference-completed observation to retrieve the best matching fuzzy belief state record that contains a decision as an extra variable. The process is analogous to case-based reasoning, but it uses fuzzification to ameliorate uncertainty and to complete missing data. The test results on real world data prove the effectiveness of this methodology.

Published by Elsevier B.V.

1. Introduction

Decisions are needed in various application domains such as pattern recognition and classification, diagnosis, prognosis, product design, marketing, military strategy, scheduling and negotiations. An outstanding problem is how to make decisions with uncertain and incomplete data from disparate sources without NP-hard algorithms. Uncertainty and incompleteness exist in every application domain. Examples are medical diagnosis where expense and timeliness cause missing data. Many decisions have to be made based on uncertain premises and incomplete information.

The uncertainty caused by noise in the available values can be severe in complex domains such as medicine, investment, aerospace, and military planning [1], due to their more sophisticated inter-variable relationships. A practical intelligent system should be able to address uncertainty and incompleteness in the data.

Existing tools for decisionmaking under uncertainty or incompleteness are either NP-hard and their parameters are very difficult and costly to obtain, or they lack the flexibility of completing missing data. Statistical methods, especially Bayesian methods, have traditionally been used for uncertainty and incompleteness but they depend on known or assumed prior distribution models and become computationally complex due to the combinatorial explosion. Dempster–Shafer Theory [2,3] is an early approach for uncertain reasoning, but its main difficulty is that we have to consider all subsets and assign probabilities. Bayesian belief networks (BBNs) [4,5] can also accommodate uncertainty and incomplete data from disparate sources. But they are NP-hard [6,7] and the necessary conditional probabilities are difficult or impossible to obtain.

Two recent fuzzy approaches, fuzzy belief networks (FBNs) [8] and fuzzy belief Petri nets (FBPNs) [9], handle uncertainty and incompleteness, avoiding not only the NP-hard computation of BBNs but also the necessity for conditional probabilities. They require a domain expert to set up network connections, as do all influence networks, which originated in 1921 [10]. Rule-based systems [1,11–13] attempt to deal with uncertainty by means of certainty factors, which are neither axiomatic nor consistent. Completeness is required by non-statistical operations research (see [14] for a connection to BBNs), mathematical models and data mining [15–17], the latter of which associates disparate data in tabular columns. Fuzzy logic methods require the completeness of

Abbreviations: FIND, fuzzy-inferenced decisionmaking; FBS, fuzzy-belief state; FBSB, fuzzy-belief-state base; FBSBR, fuzzy-belief-state-based reasoning; FSMF, fuzzy set membership function.

* Corresponding author.
E-mail addresses: Lliang@udc.edu (L.R. Liang), looney@cse.unr.edu (C.G. Looney), vinay_mandal@hotmail.com (V. Mandal).

1568-4946/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.asoc.2011.01.026

Nomenclature

N dimension of the system
$s_n(t)$ or s_n the n-th state variable at time t
$\boldsymbol{S}(\boldsymbol{t}) = (s_1(t), s_2(t), \ldots, s_N(t))$ state of a system at time t
V_n linguistic value of s_n (either LOW$_n$, MED$_n$ or HI$_n$)
f_n the fuzzy truth that s_n is V_n
$\boldsymbol{F}(\boldsymbol{t}) = (V_1(t), f_1;\ V_2(t), f_2;\ V_n(t), f_n; \ldots;\ V_N(t), f_N)$ fuzzy-belief state
$\boldsymbol{F}(\boldsymbol{t}) = (V_{11}{:}f_{11}, V_{12}{:}f_{12};\ V_{21}{:}f_{21}, V_{22}{:}f_{22}; \ldots;\ V_{n1}{:}f_{n1}, V_{n2}{:}f_{n2}; \ldots;\ V_{N1}{:}f_{N1}, V_{N2}{:}f_{N2})$ dual-mode fuzzy-belief state
Q number of state vectors in a collection of vectors
$\mu_{n\mathrm{L}}, \mu_{n\mathrm{M}}, \mu_{n\mathrm{H}}$ fuzzy means of LOW$_n$, MED$_n$ and HI$_n$ fuzzy set
$\sigma_{n\mathrm{L}}, \sigma_{n\mathrm{M}}, \sigma_{n\mathrm{H}}$ standard deviations of LOW$_n$, MED$_n$ and HI$_n$ fuzzy set
$F_{n\mathrm{L}}(s)$, $F_{n\mathrm{M}}(s)$, $F_{n\mathrm{H}}(s)$ fuzzy set membership functions for LOW$_n$, MED$_n$ and HI$_n$
$\boldsymbol{B}$ fuzzy-belief-state base (FBSB)
$\boldsymbol{S}^{\sim}(t_m)$ observed incomplete state at time t_m
$\{s_j\}$ observed variables
$\{s_k\}$ unobserved variables
$\boldsymbol{F}^{\sim}(t_m)$ incomplete fuzzy-belief state at time t_m
$f_{j\mathrm{L}}(t_m), f_{j\mathrm{M}}(t_m), f_{j\mathrm{H}}(t_m)$ fuzzy beliefs of s_j at t_m
$\boldsymbol{R}_1, \boldsymbol{R}_2$ sets of system variables
s_x, s_y variables in $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ respectively
$\boldsymbol{X}$ a pair of statements $\{s_{\boldsymbol{x}}$ is $V_{\boldsymbol{x}}\}$
$\boldsymbol{Y}$ a statement $\{s_{\boldsymbol{y}}$ is $V_{\boldsymbol{y}}\}$
V_x linguistic value of s_x (LOW$_x$, MED$_x$, or HI$_x$)
$\boldsymbol{X} \to \boldsymbol{Y}(b_{xy}, C_{xy})$ b_{xy} is the belief of $\boldsymbol{Y}$ when $\boldsymbol{X}$ is true; C_{xy} is the rule confidence
$f_{k\mathrm{LOW}}, f_{k\mathrm{MED}}, f_{k\mathrm{HI}}$ fuzzy beliefs that s_k is LOW$_k$, is MED$_k$, is HI$_k$
$\boldsymbol{F}'(t_m)$ inference-completed fuzzy-belief-state at time t_m
λ_q similarity of the q-th vector in the FBSB to $\boldsymbol{F}'(t_m)$
α_{qn} the vote of template matching of the q-th vector to $\boldsymbol{F}'(t_m)$ on the n-th dimension.
δ_{qn} the square root of the mean square error between the beliefs of the q-th vector and $\boldsymbol{F}'(t_m)$ on dimension n.

the input data, as do genetic algorithms. Rough sets [18] require certainty as well. Neural networks [19] generally require complete data as do pattern recognition and classification [20]. Case-based reasoning (CBR) is powerful [21–24] but requires certainty and incompleteness degrades its performance. In [25], CBR is integrated with clustering and feature selection to improve performance. Still, this approach requires completeness of data.

For the above methods, missing data need to be handled first before they can be applied. Traditional ways of handling missing data were described and experimented with in [26]. Their performances were illustrated and compared against each other. In a recent research effort [27], case-based reasoning was integrated with neural network for prediction. However, the missing data were still imputed first by statistical means.

In this paper, we introduce a new methodology to solve the stated problem, which is called *fuzzy-inferenced decisionmaking* (*FIND*). In FIND, a fuzzy-belief-state base (FBSB) is created from historical data of the states of a system by clustering the set of values for each state variable into three clusters upon whose center fuzzy set membership functions LOW, MEDIUM and HIGH are defined. The FBSB is mined for fuzzy association rules using the fuzzy set memberships to infer values for the missing data via these rules. When given an incomplete and uncertain observation of the system state, the observed state is completed via fuzzy association rules. Then each case in the FBSB is matched against the inference-

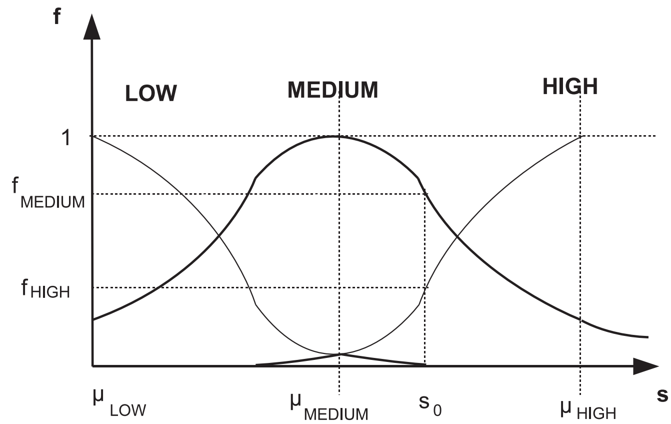

Fig. 1. Three FSMF of LOW, MEDIUM and HIGH for *x*.

completed observation to retrieve the best matching fuzzy belief state record that contains a decision as an extra variable. The process is analogous to case-based reasoning, but it uses fuzzification to ameliorate uncertainty and to complete missing data.

FIND explores the problem from a data-mining point of view, and integrates data mining with belief inferencing, case-based reasoning and machine learning. It solves the stated problem by reusing the historical information of the system in consideration and explores the implicit belief network underneath the data without the need for structural representation of graphical influences. It uses simple associations that are not NP-hard instead of combinatorial (NP-hard) conditional probabilities that Bayesian belief networks use.

Section 2 introduces the concept of fuzzy-belief-state base (FBSB) for *FIND*. Section 3 describes the first phase of *FIND* for creating a FBSB. Section 4 provides the second phase, which is a novel reasoning paradigm for fuzzy-belief-state-based reasoning. We apply *FIND* to real world data in Section 5. Time complexity of FIND is analyzed in Section 6. Section 7 concludes the paper.

2. Fuzzy-belief-state base concepts: the base of *FIND*

The *state* of a stationary system that is described by N variables is an N-dimensional *state vector* $\boldsymbol{S}(t_i) = (s_1(t_i), s_2(t_i), \ldots, s_N(t_i))$ where the components are *state variables*. We can suppress the time and write s_n in place of $s_n(t)$. A *fuzzy set D* is a contiguous set of real numbers that has a *fuzzy set membership function* (FSMF) $g(s)$ defined for all real numbers s. For every value of s the FSMF provides the *fuzzy truth* $f = g(s)$, $0 \le g(s) \le 1$, of membership of s in D. A fuzzy set is represented by a linguistic value such as LOW, MEDIUM or HIGH, or other, such as MEDIUM HIGH.

Fig. 1 shows three Gaussian FSMFs for the variable s, where the value $s = s_0$ has membership in the FSMF (called MEDIUM) of f_{MEDIUM} and also has membership in the FSMF (HIGH) of f_{HIGH}. If s represents temperature, then in the figure a value of s_0 is strongly MEDIUM but is also HIGH to a small extent. These two fuzzy truths are enough to specify this value, given the FSMFs. These FSMFs are defined below.

$$g_{\mathrm{LOW}}(s) = \exp\left[\frac{-(s-\mu_{\mathrm{LOW}})^2}{(2\sigma^2)}\right], \qquad s < \mu_{\mathrm{LOW}}, \quad \text{else } 0 \tag{1a}$$

$$g_{\mathrm{MEDIUM}}(s) = \exp\left[\frac{-(s-\mu_{\mathrm{MEDIUM}})^2}{(2\sigma^2)}\right], \quad \text{all } s \tag{1b}$$

$$g_{\mathrm{HIGH}}(s) = \exp\left[\frac{-(s-\mu_{\mathrm{HIGH}})^2}{(2\sigma^2)}\right], \qquad s > \mu_{\mathrm{HIGH}}, \quad \text{else } 0 \tag{1c}$$