Van-Nam Huynh · Masahiro Inuiguchi
Bac Le · Bao Nguyen Le
Thierry Denoeux (Eds.)

LNAI 9978

Integrated Uncertainty in Knowledge Modelling and Decision Making

5th International Symposium, IUKM 2016
Da Nang, Vietnam, November 30 – December 2, 2016
Proceedings

Springer

Lecture Notes in Artificial Intelligence 9978

Subseries of Lecture Notes in Computer Science

More information about this series at http://www.springer.com/series/1244

Van-Nam Huynh · Masahiro Inuiguchi
Bac Le · Bao Nguyen Le
Thierry Denoeux (Eds.)

Integrated Uncertainty in Knowledge Modelling and Decision Making

5th International Symposium, IUKM 2016
Da Nang, Vietnam, November 30 – December 2, 2016
Proceedings

Springer

Editors
Van-Nam Huynh
Japan Advanced Institute of Science and Technology
Nomi, Ishikawa
Japan

Masahiro Inuiguchi
Graduate School of Engineering Science
Osaka University
Toyonaka, Osaka
Japan

Bac Le
University of Science
Ho Chi Minh City
Vietnam

Bao Nguyen Le
Duy Tan University
Da Nang
Vietnam

Thierry Denoeux
Université de Technologie de Compiègne
Compiègne
France

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-319-49045-8 ISBN 978-3-319-49046-5 (eBook)
DOI 10.1007/978-3-319-49046-5

Library of Congress Control Number: 2016955999

LNCS Sublibrary: SL7 – Artificial Intelligence

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Sequential Decision Process Supported by a Compositional Model

Radim Jiroušek[1,2(✉)] and Lucie Váchová[1]

[1] Faculty of Management, University of Economics, Jindřichův Hradec, Czech Republic
radim@utia.cas.cz, vachova@fm.vse.cz
[2] Institute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czech Republic

Abstract. The goal of the paper is to describe a classical sequential decision process that is often used for both medical and technical diagnosis making in a relatively new theoretical setting. For this, we represent the background knowledge, which is assumed to be expressed in the form of a multidimensional probability distribution, as a compositional model. Though we do not perform a detailed analysis of its computational complexity, we show that the whole process is easily tractable for probability distributions of very high dimensions in the case that the distribution is represented as a compositional model of special properties.

Keywords: Sequential decision-making · Multidimensional models · Probability distributions · Composition · Information theory

1 Introduction

The basic idea, upon which a sequential decision process is based, is simple. The goal is to find out evidence supporting a decision with a required certainty. To express this idea more precisely let us use the language of probability theory. Let Z denote a decision variable whose values correspond to individual decisions (e.g., diagnoses). If κ is a probability distribution describing a relationship among the feature variables and the decision variable, we want to find out a subset E of feature variables whose values $\mathbf{e}$ (evidence) observed in the situation in question yields

$$\kappa(Z = \mathbf{d}|E = \mathbf{e}) \geq \varepsilon$$

for some decision $\mathbf{d}$, and a required reliability threshold ε. In a sequential process, the set E is gradually constructed by adding (usually) one variable at each step. Naturally, we want to keep the set E as small as possible, or generally, if some weights (or costs) connected with individual variables are given, we want to get the cheapest possible decision process.

Quite often, sequential decision processes are represented in a form of a decision tree. However, in many practical problems a fixed tree does not suit the

V.-N. Huynh et al. (Eds.): IUKM 2016, LNAI 9978, pp. 51–63, 2016.
DOI: 10.1007/978-3-319-49046-5_5

situations. For example, in case of a medical diagnosis making, a patient visits a physician with some complains, i.e., with symptoms (values of variables) that need not be at the beginning of the decision tree, and still they should be included in the set E. Similarly, faults of a technical device manifest at the very beginning in different ways (or, it may be observed by different users in different ways). Not speaking about a newly appointed manager whose task is to strengthen a company. To diagnose weaknesses of the enterprise they always start with different prior knowledge. Moreover, costs assigned to variables may differ from case to case. A patient may come to a specialist with the results of some expensive laboratory tests that were already performed by another physician, so the specialist has it free. This is why we prefer not to compute a single decision tree in advance but to construct a sequential decision process based on a multidimensional probability distribution at time of its application.

When speaking about dimensionality of probability distributions, it is clear that in practical situations we should have in mind rather thousands than tens of variables. Therefore we will use a tool enabling us to represent and compute with such large distributions.

For this purpose, Bayesian networks [5] that were developed in 1980s, or other graphical models [10] are often used. In contrast to this, in this paper we want to enhance application of another, non-graphical approach, so called compositional models that were proposed for multidimensional probability distributions; see [6].

The paper is organized as follows: Sect. 2 introduces a necessary notation and serves as a brief introduction to compositional models (more on this topic can be found in [6]). In Sect. 3 we show that it is possible to compute conditionals even for probability distributions of very high dimensions when the distributions are represented in the form of compositional models of special properties. Section 4 is devoted to the description of the sequential decision process.

2 Compositional Model

In this text we use the notation from the paper [7] presented at IUKM 2015. We deal with a finite system of finite-valued random variables $N = \{X_1, \ldots, X_n, Z\}$. A set of values of feature variable X_i, which is denoted $\mathbb{X}_i$, is assumed to have at least two elements. The same is assumed also about set $\mathbb{Z}$ of values of decision variable Z. The set of all combinations of the considered values is denoted $\mathbb{X}_N = \mathbb{X}_1 \times \mathbb{X}_2 \times \ldots \times \mathbb{X}_n \times \mathbb{Z}$. Analogously, for $K \subset N \setminus \{Z\}$, $\mathbb{X}_K = \times_{X_i \in K} \mathbb{X}_i$, and for $Z \in K \subset N$, $\mathbb{X}_K = \times_{X_i \in K} \mathbb{X}_i \times \mathbb{Z}$.

Distributions of the considered variables are denoted by Greek letters (π, κ, μ) with possible indices; thus for $K \subseteq N$, we can consider a distribution $\pi(K)$, which is a $|K|$-dimensional distribution and $\pi(x)$ denotes a value of probability distribution π for state $x \in \mathbb{X}_K$.

For a probability distribution $\pi(K)$ and $J \subset K$, its *marginal distribution* $\pi^{\downarrow J}$ is computed for all $x \in \mathbb{X}_J$ by

$$\pi^{\downarrow J}(x) = \sum_{y \in \mathbb{X}_K : y^{\downarrow J} = x} \pi(y),$$

where $y^{\downarrow J}$ denotes the *projection* of $y \in \mathbb{X}_K$ into $\mathbb{X}_J$. For computation of marginal distributions we do not exclude situations when $J = \emptyset$, in this case $\pi^{\downarrow\emptyset} = 1$.

Having two distributions $\pi(K)$ and $\kappa(K)$, we say that κ dominates π (in symbol $\pi \ll \kappa$) if for all $x \in \mathbb{X}_K$

$$\kappa(x) = 0 \implies \pi(x) = 0.$$

One of the most important notions supporting an efficient representation of multidimensional probability distributions is a famous concept of conditional independence (se e.g. [10] or [13]).

For a probability distribution $\pi(K)$ and three disjoint subsets $L, M, R \subseteq K$ such that both $L, M \neq \emptyset$, we say that groups of variables L and M are *conditionally independent* given R (in symbol $L \perp\!\!\!\perp M|R\ [\pi]$) if

$$\pi^{\downarrow L\cup M\cup R}\pi^{\downarrow R} = \pi^{\downarrow L\cup R}\pi^{\downarrow M\cup R}.$$

2.1 Operator of Composition

For the compositional models, the key notion is that of a composition.

Definition 1. *For arbitrary distributions $\pi(K)$ and $\kappa(L)$, for which $\pi^{\downarrow K\cap L} \ll \kappa^{\downarrow K\cap L}$, their* composition *is, for each $x \in \mathbb{X}_{(L\cup K)}$, given by the following formula*

$$(\pi \triangleright \kappa)(x) = \frac{\pi(x^{\downarrow K})\kappa(x^{\downarrow L})}{\kappa^{\downarrow K\cap L}(x^{\downarrow K\cap L})}.$$

In case $\pi^{\downarrow K\cap L} \not\ll \kappa^{\downarrow K\cap L}$, the composition remains undefined.

Let us briefly repeat the basic properties of this operator that were discussed in more details in [6] and also in [7].

Lemma 1. *Suppose $\pi(K)$, $\kappa(L)$ and $\mu(M)$ are such probability distributions that all the following expressions (compositions) are defined. Then the following statements hold.*

1. *$\pi \triangleright \kappa$ is a probability distribution of variables $(L \cup K)$ and its marginal distribution for variables K equals π:*

$$(\pi \triangleright \kappa)^{\downarrow K} = \pi.$$

2. *The composition is not commutative, therefore in general: $\pi \triangleright \kappa \neq \kappa \triangleright \pi$, however, π and κ are consistent, i.e., $\pi^{\downarrow K\cap L} = \kappa^{\downarrow K\cap L}$, if and only if*

$$\pi \triangleright \kappa = \kappa \triangleright \pi.$$

3. *The composition is not associative, therefore in general $(\pi \triangleright \kappa) \triangleright \mu \neq \pi \triangleright (\kappa \triangleright \mu)$, however if $K \supset (L \cap M)$, or, $L \supset (K \cap M)$, then*

$$(\pi \triangleright \kappa) \triangleright \mu = \pi \triangleright (\kappa \triangleright \mu).$$

4. If $(K \cap L) \subseteq R \subseteq K \cup L$*, then*

$$(\pi \triangleright \kappa)^{\downarrow R} = \pi^{\downarrow K \cap R} \triangleright \kappa^{\downarrow L \cap R}.$$

5. If $K \supset (L \cap M)$*, then* $(\pi \triangleright \kappa) \triangleright \mu = (\pi \triangleright \mu) \triangleright \kappa$*.*

The above presented assertion expresses all the properties of the operator of composition we will need in this paper. For example, from Property 1 one can easily see that it enables us to construct a more-dimensional distribution from two low-dimensional ones. The result is an *extension* of the first argument. On the other hand side, having a more-dimensional distribution one can, in some situations, factorize this distribution into two (or more) low-dimensional ones. This property is precisely expressed in the following assertion, the proof of which can also be found in [6].

Lemma 2. *Consider a probability distribution* $\pi(K)$*, and three disjoint subsets* $L, M, R \subseteq K$ *such that both* $L, M \neq \emptyset$*. Then* $L \perp\!\!\!\perp M | R$ $[\pi]$ *if and only if*

$$\pi^{\downarrow L \cup M \cup R} = \pi^{\downarrow L \cup R} \triangleright \pi^{\downarrow M \cup R}.$$

2.2 Generating Sequences

In the rest of the paper we will deal with sequences of low-dimensional probability distributions. To avoid some technical problems and the necessity of repeating some assumptions to excess, let us make the following three conventions.

First, whenever we speak about a distribution π_k, it will be a distribution $\pi_k(K_k)$. Thus, formula $\pi_1 \triangleright \ldots \triangleright \pi_m$, if it is defined, determines the distributions of variables $K_1 \cup \ldots \cup K_m$.

Since the operator of composition is not associative (see Property 3 of Lemma 1), the formulas like $\pi_1 \triangleright \ldots \triangleright \pi_m$ are, strictly speaking, ambiguous. Therefore, the second convention avoids this ambiguity saying that we always apply the operators from left to right. Thus

$$\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \ldots \triangleright \pi_m = (((\pi_1 \triangleright \pi_2) \triangleright \pi_3) \triangleright \ldots \triangleright \pi_m),$$

and the parentheses will be used only when we want to change this default ordering. Therefore, to construct a multidimensional distribution it is sufficient to determine a sequence – we call it a *generating sequence* – of low-dimensional distributions.

The third aforementioned convention is of a rather technical nature. Considering a generating sequence $\pi_1, \pi_2, \ldots, \pi_m$ we will always assume that all the included operators of composition are well defined, which means that we assume

$$\pi_1^{\downarrow K_2 \cap K_1} \ll \pi_2^{\downarrow K_2 \cap K_1},$$
$$(\pi_1 \triangleright \pi_2)^{\downarrow K_3 \cap (K_1 \cup K_2)} \ll \pi_3^{\downarrow K_3 \cap (K_1 \cup K_2)},$$

and so on. Therefore, we assume that $\pi_1 \triangleright \pi_2 \triangleright \pi_3 \triangleright \ldots \triangleright \pi_m$ is defined.

Another important notion we will need in this paper is that of a perfect sequences that is defined in the following way (see also [6]).

Definition 2. *A generating sequence of probability distributions* $\pi_1, \pi_2, \ldots, \pi_m$ *is called* perfect *if*

$$\begin{aligned}
&\pi_1 \triangleright \pi_2 = \pi_2 \triangleright \pi_1,\\
&\pi_1 \triangleright \pi_2 \triangleright \pi_3 = \pi_3 \triangleright (\pi_1 \triangleright \pi_2),\\
&\quad\vdots\\
&\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_m = \pi_m \triangleright (\pi_1 \triangleright \ldots \triangleright \pi_{m-1}).
\end{aligned}$$

The importance of perfect sequences becomes clear from the following characterization theorem (for the proofs of all the assertions presented in the rest of this section see [6]).

Theorem 1. *A sequence of distributions* $\pi_1, \pi_2, \ldots, \pi_m$ *is perfect* iff *all the distributions from this sequence are marginals of the distribution* $\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_m$.

Another advantageous property of the perfect sequences says that perfect sequences represent a unique distribution in the following sense.

Theorem 2. *If a sequence* $\pi_1, \pi_2, \ldots, \pi_m$ *and its permutation* $\pi_{i_1}, \pi_{i_2}, \ldots, \pi_{i_m}$ *are both perfect, then*

$$\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_m = \pi_{i_1} \triangleright \pi_{i_2} \triangleright \ldots \triangleright \pi_{i_m}.$$

In the rest of the paper we will need two more facts expressed in the following assertions. The first says that any generating sequence can be transformed into a perfect sequence without influencing the resulting multidimensional distribution.

Theorem 3. *For any generating sequence* $\pi_1, \pi_2, \ldots, \pi_m$, *the sequence* $\kappa_1, \kappa_2, \ldots, \kappa_m$ *computed by the following process*

$$\begin{aligned}
&\kappa_1 = \pi_1,\\
&\kappa_2 = \kappa_1^{\downarrow K_2 \cap K_1} \triangleright \pi_2,\\
&\kappa_3 = (\kappa_1 \triangleright \kappa_2)^{\downarrow K_3 \cap (K_1 \cup K_2)} \triangleright \pi_3,\\
&\quad\vdots\\
&\kappa_m = (\kappa_1 \triangleright \ldots \triangleright \kappa_{m-1})^{\downarrow K_m \cap (K_1 \cup \ldots \cup K_{m-1})} \triangleright \pi_m
\end{aligned}$$

is perfect, and

$$\pi_1 \triangleright \ldots \triangleright \pi_m = \kappa_1 \triangleright \ldots \triangleright \kappa_m.$$

The process of perfectization described in Theorem 3 is simple. Unfortunately, not from the point of view of computational complexity. This is because the computation of $(\kappa_1 \triangleright \ldots \triangleright \kappa_{r-1})^{\downarrow K_r \cap (K_1 \cup \ldots \cup K_{r-1})}$ may be computationally very expensive [2]. Therefore, in the next sections we will take advantage of special generating sequences, namely those whose sequences of variables $K_1, K_2, \ldots, K_m$ meet the so called *running intersection property* (RIP):

$$\forall r = 2, \ldots, m \quad \exists s (1 \leq s < r) \quad \left(K_r \cap (\bigcup_{k=1}^{r-1} K_k) \subseteq K_s \right).$$

Consider a perfectization process applied to such a RIP generating sequence $\pi_1, \pi_2, \ldots, \pi_m$. In this case, either $K_1 \supseteq K_3 \cap (K_1 \cup K_2)$, or $K_2 \supseteq K_3 \cap (K_1 \cup K_2)$, which means that $K_3 \cap (K_1 \cup K_2)$ equals either $K_1 \cap K_3$ or $K_2 \cap K_3$. Since the sequence $\kappa_1, \ldots, \kappa_m$ is constructed in the way that κ_1 and κ_2 are consistent, i.e., $\kappa_1^{\downarrow K_2 \cap K_1} = \kappa_2^{\downarrow K_2 \cap K_1}$, due to Property 2 of Lemma 1 $\kappa_1 \triangleright \kappa_2 = \kappa_2 \triangleright \kappa_1$ and therefore, using Property 1 of the same Lemma, $(\kappa_1 \triangleright \kappa_2)^{\downarrow K_1} = \kappa_1$ and $(\kappa_1 \triangleright \kappa_2)^{\downarrow K_2} = \kappa_2$. Thus we got that $(\kappa_1 \triangleright \kappa_2)^{\downarrow K_3 \cap (K_1 \cup K_2)}$ equals either $\kappa_1^{\downarrow K_3 \cap K_1}$, or $\kappa_2^{\downarrow K_3 \cap K_2}$. Analogously, we can see that for any $r = 3, \ldots, m$ there exists s such that

$$(\kappa_1 \triangleright \ldots \triangleright \kappa_{r-1})^{\downarrow K_r \cap (K_1 \cup \ldots \cup K_{r-1})} = \kappa_s^{\downarrow K_r \cap K_s},$$

which makes the perfectization process computationally simple.

Another advantageous property concerning the RIP generating sequences is expressed in the following theorem [6].

Theorem 4. *If $\pi_1, \ldots, \pi_m$ is a sequence of pairwise consistent probability distributions such that $K_1, \ldots, K_m$ meets RIP, then this sequence is perfect.*

3 Conditioning

In [3] we showed that even the conditional distribution can be computed using the operator of composition. Consider any variable $Y \in N$ and its value $\mathbf{a} \in \mathbb{X}_{\{Y\}}$. Define a *degenerated* one-dimensional probability distribution $\delta_{\mathbf{a}}(Y)$ as a distribution of variable Y achieving probability 1 for value $Y = \mathbf{a}$, i.e.,

$$\delta_{\mathbf{a}}(Y) = \begin{cases} 1 & \text{if } Y = \mathbf{a}, \\ 0 & \text{otherwise.} \end{cases}$$

Now, compute $(\delta_{\mathbf{a}}(Y) \triangleright \kappa)^{\downarrow \{X\}}$ for a probability distribution $\kappa(L)$ with $X, Y \in L$. For $\mathbf{b} \in \mathbb{X}_{\{X\}}$

$$\begin{aligned}(\delta_{\mathbf{a}}(Y) \triangleright \kappa)^{\downarrow \{X\}}(\mathbf{b}) &= ((\delta_{\mathbf{a}}(Y) \triangleright \kappa)^{\downarrow \{X,Y\}})^{\downarrow \{X\}}(\mathbf{b}) = (\delta_{\mathbf{a}}(Y) \triangleright \kappa^{\downarrow \{X,Y\}})^{\downarrow \{X\}}(\mathbf{b}) \\ &= \sum_{x \in \mathbb{X}_{\{Y\}}} (\delta_{\mathbf{a}}(Y) \triangleright \kappa^{\downarrow \{X,Y\}})(\mathbf{b}, x) \\ &= \sum_{x \in \mathbb{X}_{\{Y\}}} \frac{\delta_{\mathbf{a}}(Y) \cdot \kappa^{\downarrow \{X,Y\}}(\mathbf{b}, x)}{\kappa^{\downarrow \{Y\}}(x)} \\ &= \frac{\kappa^{\downarrow \{X,Y\}}(\mathbf{b}, \mathbf{a})}{\kappa^{\downarrow \{Y\}}(\mathbf{a})} = \kappa^{\downarrow \{X,Y\}}(\mathbf{b}|\mathbf{a}).\end{aligned}$$

Analogously, we can easily show that for $\kappa(L)$

$$\kappa(L \setminus \{Y\}|Y = \mathbf{a}) = (\delta_{\mathbf{a}}(Y) \triangleright \kappa)^{\downarrow L \setminus \{Y\}},$$

and for $E = \{Y_1, Y_2, \ldots, Y_k\}$, and $\mathbf{e} \in \mathbb{X}_E$,

$$\kappa(L \setminus E|E = \mathbf{e}) = (\delta_{\mathbf{e}^{\downarrow \{Y_1\}}}(Y_1) \triangleright (\ldots \triangleright (\delta_{\mathbf{e}^{\downarrow \{Y_k\}}}(Y_k) \triangleright \kappa)))^{\downarrow L \setminus E}.$$

Now, we want to show that the last expression can be effectively computed for a multidimensional distribution κ represented in a form of a RIP perfect generating sequence: $\kappa = \pi_1(K_1) \triangleright \pi_2(K_2) \triangleright \ldots \triangleright \pi_m(K_m)$.

Let us, first, consider the computation

$$\delta_{\mathbf{e}^{\downarrow\{Y_k\}}}(Y_k) \triangleright \kappa = \delta_{\mathbf{e}^{\downarrow\{Y_k\}}}(Y_k) \triangleright (\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_m).$$

It is well known (and the reader can see it from the properties of joint trees described in Sect. 4.1) that for each K_i the RIP sequence $K_1, K_2, \ldots, K_m$ can be reordered so that the resulting sequence is also RIP and K_i is at the beginning of this new sequence. So, without loss of generality we can assume that $\pi_1, \pi_2, \ldots, \pi_m$ is such that $Y_k \in K_1$. The fact that the respective reordering does not influence the represented multidimensional distribution $\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_m$ follows from Theorems 4 and 2.

Now, consider

$$\delta_{\mathbf{e}^{\downarrow\{Y_k\}}}(Y_k) \triangleright ((\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_{m-1}) \triangleright \pi_m).$$

Since Y_k is contained in K_1, it is contained also in $(K_1 \cup K_2 \cup \ldots \cup K_{m-1})$, and therefore, applying Property 3 of Lemma 1, we get

$$\begin{aligned}\delta_{\mathbf{e}^{\downarrow\{Y_k\}}}(Y_k) \triangleright ((\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_{m-1}) \triangleright \pi_m) \\ = (\delta_{\mathbf{e}^{\downarrow\{Y_k\}}}(Y_k) \triangleright ((\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_{m-1})) \triangleright \pi_m.\end{aligned}$$

Repeating this reasoning $m-1$ times we eventually get

$$\delta_{\mathbf{e}^{\downarrow\{Y_k\}}}(Y_k) \triangleright (\pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_m) = (\delta_{\mathbf{e}^{\downarrow\{Y_k\}}}(Y_k) \triangleright \pi_1) \triangleright \pi_2 \triangleright \ldots \triangleright \pi_m,$$

which means that computing a conditional from a distribution represented by a RIP perfect generating sequence results, again, in a distribution represented as a RIP sequence. The latter can be, as showed in the preceding section, efficiently transformed into a RIP perfect generating sequence (Theorem 3). Therefore, to compute $\kappa(L \setminus E | E = \mathbf{e})$ we have to successively apply the above idea for all $Y_j \in E$. Let us stress that for each Y_j we have to find a RIP ordering of $(K_1, K_2, \ldots, K_m)$ such that Y_j is in the first set of this RIP ordering. Generally, for different Y_j we have to use a different RIP ordering of $(K_1, K_2, \ldots, K_m)$.

4 Sequential Decision Process

A sequential decision process consists in a successive repetition of the following step:

Knowing state $\mathbf{e} \in \mathbb{X}_E$ *of variables* E, *find a variable* $X \in N \setminus (E \cup \{Z\})$ *such that the detection of its value* $\mathbf{a}$ *increases (as much as possible) the chances of getting*

$$\kappa(Z = \mathbf{d} | E = \mathbf{e}, X = \mathbf{a}) \geq \varepsilon$$

for some value $\mathbf{d} \in \mathbb{Z}$.

It is important to realize that in this step we search for a variable X whose value is to be ascertained next. It means that at the moment of looking for X we do not know the value $\mathbf{a} \in \mathbb{X}_{\{X\}}$. This value will be ascertain before the next sequential step is realized with new $E := E \cup \{X\}$.

For the selection of $X \in N \setminus (E \cup \{Z\})$ we can hardly find a better criterion than that used in the process of construction of efficient decision (or search) trees for many years [9,12]. Using this criterion, in the process of data based construction of a decision tree, we look for the variable that splits the considered training data file into subfiles yielding the minimum expected Shannon entropy for the decision variable. In fact, it is nothing else than looking for a variable $X \in N \setminus (E \cup \{Z\})$ maximizing the expression

$$MI_{\kappa(N\setminus E|E=\mathbf{e})}(X;Z)$$
$$= \sum_{(x,z)\in\mathbb{X}_{\{X\}}\times\mathbb{Z}} \kappa^{\downarrow\{X,Z\}}(x,z|E=\mathbf{e}) \cdot \log \frac{\kappa^{\downarrow\{X,Z\}}(x,z|E=\mathbf{e})}{\kappa^{\downarrow\{X\}}(x|E=\mathbf{e})\kappa^{\downarrow\{Z\}}(z|E=\mathbf{e})}.$$

Now, as the reader certainly expects, we take advantage of the results from the preceding section and assume that the conditional distribution $\kappa(N \setminus E|E = \mathbf{e})$ is represented in the form of a RIP perfect generating sequence. This assumption enables us not only to compute values $MI_{\kappa(N\setminus E|E=\mathbf{e})}(X;Z)$ for all $X \in N \setminus (E \cup \{Z\})$ in an efficient way, but also to speed up the computational process by indicating those variables, for which we need not to compute the value of mutual information because we can learn in advance that the respective conditional mutual information cannot achieve a maximal value.

4.1 Computations in Joint Trees

It is known from both data-base [1] and Bayesian network [5,10] theories that a system of sets $K_1, K_2, \ldots, K_m$ can be ordered to meet RIP if and only if one can construct a structure called a *joint tree*. It is a tree having $K_1, K_2, \ldots, K_m$ for its nodes and possessing the following special property: If K_k lies on the path from K_r to K_s then $K_k \supseteq K_r \cap K_s$.

Recall that it is an easy task to construct a joint tree for a RIP sequence $K_1, K_2, \ldots, K_m$: For, each K_r $(r = 2, \ldots, m)$ the joint tree contains an edge connecting K_r with that K_s, which meets the RIP condition

$$1 \leq s < r \quad \& \quad K_r \cap (K_1 \cup \ldots \cup K_{r-1}) \subseteq K_s.$$

If there are more such nodes K_s, then K_r is connected to only one of them. The tree contains no other edges than those specified above.

To compute $MI_{\kappa(N\setminus E|E=\mathbf{e})}(X;Z)$ for all $X \in N\setminus(E\cup\{Z\})$, start enumerating this mutual information for all variables from K_r, for which $Z \in K_r$ (if there are more sets meeting this condition, consider all of them). Since we assume that

$$\kappa(N \setminus E|E = \mathbf{e}) = \pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_m,$$

and that $\pi_1, \pi_2, \ldots, \pi_m$ is a perfect sequence, due to Theorem 1 we know that π_r is marginal of $\kappa(N \setminus E | E = \mathbf{e})$. This means that we can compute the required conditional mutual information just from π_r, which is simple.

After having evaluated the required mutual information for all the variables from $K_r \setminus (E \cup \{Z\})$, we start processing variables from the neighboring nodes, i.e., from nodes K_s that are adjacent to K_r in the joint tree. Now, it is important to realize two facts that makes the computation very efficient.

First, since K_s is adjacent to K_r in the joint tree, it is possible to find a RIP ordering of $K_1, K_2, \ldots, K_m$ such that it starts $K_r, K_s, ...$, and therefore (thanks to Theorem 2 and Property 1 of Lemma 1) we know that

$$\kappa((K_r \cup K_s) \setminus E | E = \mathbf{e}) = \pi_r \triangleright \pi_s.$$

The other fact that can speed up the computational process follows from a famous property of mutual information. It is known from any textbook on information theory (e.g. [4]) that if $X \perp\!\!\!\perp Z | M[\mu]$, for some set of variables M then

$$MI_\mu(X; Z) \leq MI_\mu(M; Z).$$

Therefore, applying this property and Lemma 2 to this situation we get that for $X \in K_s \setminus K_r$ (recall that $Z \in K_r \setminus K_s$ because all variables from K_r were treated in the previous step)

$$MI_{\pi_r \triangleright \pi_s}(X; Z) \leq MI_{\pi_r \triangleright \pi_s}(K_r \cap K_s; Z).$$

This means that if there is variable $X \in K_r$ for which

$$MI_{\pi_r \triangleright \pi_s}(K_r \cap K_s; Z) \leq MI_{\pi_r \triangleright \pi_s}(X; Z)$$

then we do not need to compute mutual information $MI_{\pi_r \triangleright \pi_s}(X'; Z)$ for variables $X' \in K_s \setminus K_r$ because we know that it cannot achieve the looked for maximum.

In a similar way we can compute $MI_{\kappa(N \setminus E | E = \mathbf{e})}(X; Z)$ for all the remaining variables from $N \setminus (E \cup \{Z\})$. First we have to find the shortest path in the considered joint tree connecting nodes containing X and Z (realize that in a tree, two nodes are always connected by a unique path, but both the considered variables X and Z may be in several nodes). Denote this path $K_{j_1}, K_{j_2}, \ldots, K_{j_k}$ in the way that $X \in K_{j_1}$ and $Z \in K_{j_k}$. Then from the perfectness of $\pi_1, \pi_2, \ldots, \pi_m$ and the RIP property we get that

$$\kappa((K_{j_1} \cup K_{j_2} \cup \ldots \cup K_{j_k}) \setminus E | E = e) = \pi_{j_1} \triangleright \pi_{j_2} \triangleright \ldots \triangleright \pi_{j_k},$$

and thus we compute

$$MI_{\kappa(N \setminus E | E = \mathbf{e})}(X; Z) = MI_{\pi_{j_1} \triangleright \pi_{j_2} \triangleright \ldots \triangleright \pi_{j_k}}(X; Z).$$

However, the longer path $K_{j_1}, K_{j_2}, \ldots, K_{j_k}$ the greater chances that among the variables, for which the mutual information has been evaluated, we can find variable X', for which

$$MI_{\pi_{j_1} \triangleright \pi_{j_2} \triangleright \ldots \triangleright \pi_{j_k}}(K_{j_i} \cap K_{j_{i+1}}; Z) \leq MI_{\pi_{j_1} \triangleright \pi_{j_2} \triangleright \ldots \triangleright \pi_{j_k}}(X'; Z),$$

for some $i \in \{1, 2, k - 1\}$, and therefore the computation of $MI_{\pi_{j_1} \triangleright \pi_{j_2} \triangleright \ldots \triangleright \pi_{j_k}}(X; Z)$ is wasteful.

4.2 An Algorithm

Up to now, we have described and theoretically supported all the individual parts of the proposed sequential decision process. Let us (rather informally) summarize it in several steps.

Initialization. We assume that the probability distribution represented by a RIP generating sequence is given. If it is not perfect then apply the perfectization procedure described in Theorem 3 so that the resulting generating sequence $\pi_1, \pi_2, \ldots, \pi_m$ meets RIP condition and is perfect. Thus, in the sequel we compute with the distribution

$$\kappa(N) = \pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_m.$$

Set $E := \emptyset$.

Preliminary Evidence Processing. E_0 denotes the set of variables whose values $\mathbf{e}_0$ are given before the sequential process starts. If there is no preliminary evidence, i.e. $E_0 = \emptyset$, skip the rest of this step. Otherwise for all variables X from E_0 realize the following conditioning procedure.

Conditioning. Assign a new value to $E := E \cup \{X\}$, and extend the point $\mathbf{e} := \mathbf{e}_0^{\downarrow E}$. Reorder (renumber) the given generating sequence in the way that the new ordering $K_1, K_2, \ldots, K_m$ meets RIP and $X \in K_1$. Apply the perfectization procedure (Theorem 3) to the sequence: $(\delta_{\mathbf{e}_0^{\downarrow X}}(X) \triangleright \pi_1), \pi_2, \ldots, \pi_m$, and assign the result, after marginalizing variable X out, as a new value to $\pi_1, \pi_2, \ldots, \pi_m$. So that now,

$$\kappa(N \setminus E | E = \mathbf{e}) = \pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_m.$$

Sequential Procedure. Perform the following process consisting of three steps (**Variable selection**, **Application** and **Conditioning**) repeatedly until $\pi_r^{\downarrow\{Z\}}(\mathbf{d}) \geq \varepsilon$ for some $\mathbf{d} \in \mathbb{Z}$. (Take any r such that $Z \in K_r$.)

Variable selection. Set $L := N \setminus (E \cup \{Z\})$ (L is a set of variables, for which we should compute the value of mutual information in this step).
For all $X \in L$, for which there exists $r \in \{1, 2, \ldots, m\}$ such that both $X, Z \in K_r$, compute $MI_{\pi_r}(X; Z)$, and reset $L := L \setminus \{X\}$.

Now, denote X the variable achieving the maximal value of the mutual information, and repeat the following step until $L = \emptyset$.

Computation in a joint tree. Find the shortest path $K_{j_1}, K_{j_2}, \ldots, K_{j_k}$ from the respective joint tree meeting the following three properties: (1) $Z \in K_{j_k}$, (2) $K_{j_1} \cap L \neq \emptyset$, and (3) $(K_{j_2} \cup \ldots \cup K_{j_k}) \cap L = \emptyset$.
If

$$MI_{\pi_{j_2} \triangleright \ldots \triangleright \pi_{j_k}}(K_{j_1} \cap K_{j_2}; Z) > MI_{\pi_{j_2} \triangleright \ldots \triangleright \pi_{j_k}}(X; Z)$$

Then for all $Y \in K_{j_1} \cap L$ compute

$$MI_{\pi_{j_1} \triangleright \ldots \triangleright \pi_{j_k}}(Y; Z),$$

and reset $L := L \setminus K_{j_1}$.
If the maximum from the computed values of mutual information is higher than

$$MI_{\pi_{j_2} \triangleright \ldots \triangleright \pi_{j_k}}(X; Z)$$

then reset X to the variable with the highest mutual information.
Else reset L by removing from L all such K_s for which the path from K_s to K_{j_k} goes through K_{j_1}.

Application. Ask the user to ascertain the value of variable X. Denote the result **a**.

Conditioning. Reset $E := E \cup \{X_i\}$, and extend the point **e** so that the new value is from $\mathbb{X}_E$, and $e^{\downarrow\{X\}} = \mathbf{a}$. Reorder (renumber) generating sequence $\pi_1, \pi_2, \ldots, \pi_m$ in the way that the respective new ordering $K_1, K_2, \ldots, K_m$ meets RIP and $X \in K_1$. Apply the perfectization procedure (Theorem 3) to the sequence: $(\delta_{\mathbf{a}}(X) \triangleright \pi_1), \pi_2, \ldots, \pi_m$, and the result, after marginalizing variable X out, assign as a new value to $\pi_1, \pi_2, \ldots, \pi_m$. So that now,

$$\kappa(N \setminus E | E = \mathbf{e}) = \pi_1 \triangleright \pi_2 \triangleright \ldots \triangleright \pi_m.$$

5 Summary and Conclusions

We have described a sequential diagnosis making process based on a knowledge represented by a multidimensional probability distribution. Unfortunately, because of the restrictions on number of pages we are not able to include an illustrative example that will be presented in the conference lecture. Nevertheless, the idea of the process is simple and can be easily understood from the algorithm description. The reader certainly realized that the controlling rule aims for the least number of variables whose values are to be ascertain. As said in Introduction, it is really not difficult to modify the variable selection rule so that some weights of the variables are taken into account. In this case, however, one can hardly rely upon the fact that the longer path $K_{j_1}, K_{j_2}, \ldots, K_{j_k}$ is constructed in the **Computation in a joint tree** step the greater chances to cut off the nodes of the joint tree that cannot contain a variable optimizing the selection criterion.

Applying just basic properties of the operator of composition and perfect generating sequences recalled in Sect. 2 we showed that all the necessary computations, including the computation of the required conditionals, can be performed *locally*, and therefore very efficiently. Application of local computational procedures, based on original Lauritzen and Spiegelhalter ideas [11], was made possible due to specific properties of perfect RIP generating sequences.

As it can be seen from the previous text, it is the very application of perfect compositional models that have one great advantage visible in comparison with

application of Bayesian networks. Each low-dimensional distribution, which the model is constructed from, is marginal to the multidimensional model. This makes verification of some conditions, like for example the stoping rule from the Algorithm, very simple. More generally, there is a whole class of sets of variables, for which the marginals can easily be computed. These are the sets that can be got as a union of nodes of a subtree (connected subgraph) of the respective joint tree. This property was exploited in the **Computation in a joint tree** step of the Algorithm, where we considered marginals $\pi_{j_1} \triangleright \ldots \triangleright \pi_{j_k}$.

Though it is beyond the scope of this paper, let us mention yet another advantage of the considered compositional models. The operator of composition was introduced also in possibility theory [14] and recently even in Shenoy's Valuation Based Systems [8], which, as a generic uncertainty calculus covers many other calculi, such as Spohn's epistemic belief theory, and D-S belief function theory. This makes it possible to apply compositional models, and all the methods based on the compositional models like the described sequential decision process, in all these alternative uncertainty theories (naturally under the assumption that we have a function with the properties of mutual information at our disposal).

Acknowledgement. This research was partially supported by GAČR under Grant No. 15-00215S.

References

1. Beeri, C., Fagin, R., Maier, D., Yannakakis, M.: On the desirability of acyclic database schemes. J. ACM **30**(3), 479–513 (1983)
2. Bína, V., Jiroušek, R.: Marginalization in multidimensional compositional models. Kybernetika **42**(4), 405–422 (2006)
3. Bína, V., Jiroušek, R.: On computations with causal compositional models. Kybernetika **51**(3), 525–539 (2015)
4. Csiszár, I., Korner, J.: Information Theory: Coding Theorems for Discrete Memoryless Systems. Akademiai Kiado, New York (1997)
5. Jensen, F.V.: Bayesian Networks and Decision Graphs. IEEE Computer Society Press, New York (2001)
6. Jiroušek, R.: Foundations of compositional model theory. Int. J. Gen. Syst. **40**(6), 623–678 (2011)
7. Jiroušek, R., Krejčová, I.: Minimum description length principle for compositional model learning. In: Huynh, V.-N., Inuiguchi, M., Denoeux, T. (eds.) IUKM 2015. LNCS (LNAI), vol. 9376, pp. 254–266. Springer, Heidelberg (2015). doi:10.1007/978-3-319-25135-6_25
8. Jiroušek, R., Shenoy, P.P.: Compositional models in valuation-based systems. Int. J. Approx. Reasoning **53**(8), 1155–1167 (2012)
9. Landa, L.N.: Opit primenenija matematitcheskoj logiki i teorii informacii k nekotorym problemam obutchenija. Voprosy psichologii **8**, 19–40 (1962)
10. Lauritzen, S.L.: Graphical models. Oxford University Press, Oxford (1996)
11. Lauritzen, S.L., Spiegelhalter, D.J.: Local computations with probabilities on graphical structures and their application to expert systems. J. Royal Stat. Soc. Series B (Methodological) **50**, 157–224 (1988)

12. Quinlan, J.R.: Induction of Decision Trees. Mach. Learn. **1**(1), 81–106 (1986)
13. Studený, M.: Probabilistic Conditional Independence Structures. Springer, Heidelberg (2005)
14. Vejnarová, J.: Possibilistic independence and operators of composition of possibility measures. In: Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, pp. 575–580 (1998)