KYBERNETIKA — VOLUME *43* (2007), NUMBER 5, PAGES 633 – 647

COMPARING ALGORITHMS BASED ON MARGINAL PROBLEM

Otakar Kříž

The paper deals with practical aspects of decision making under uncertainty on finite sets. The model is based on *marginal problem.* Numerical behaviour of 10 different algorithms is compared in form of a study case on the data from the field of rheumatology. (Five of the algorithms types were suggested by A. Perez.) The algorithms (expert systems, inference engines) are studied in different *situations* (combinations of parameters).

Keywords: graphical probabilistic models, probabilistic inference, marginal problem

AMS Subject Classification: 68T37, 62E15

1. PREFACE – MEMORIES

This paper is, similarly as the whole special issue of Kybernetika dedicated to the memory of Albert Perez and partially, it can be considered as paying off a certain debt I feel to owe him. (By the way, this feeling is probably shared by many others who were lucky to meet him professionally.) Though primarily interested in theory, A. Perez always held the idea that a good theory should be applicable in an interesting problem area. This brought him to close cooperation with people dealing with biomedical diagnostics. Even when retired, in the age of eighty, he returned to his favorite topic from the last period of active work and suggested four new algorithms for decision making based on marginal problem. The central theme of his endeavor was looking for a suitable approximation of an "all-explaining" probabilistic distribution. This was not a purposeless but a necessary step in searching for universally applicable methods for decision making. Naturally, he was interested in speed of convergence of the algorithms and in their efficiency in typical examples and therefore he made me code the basic versions of algorithms and test their behaviour. This feedback helped him also to find out situations when the algorithms behaved surprisingly and to improve insight in the problem so that better refined clones of algorithms with parameterizing could have been synthesized.

2. INTRODUCTION

In a broader sense, the paper deals with the methodology for testing performance of different decision making algorithms (expert systems, procedures) designed to solve the diagnostic problem in probabilistic context on finite sets. The "diagnostic" terminology is used just to ease up the orientation in semantics of different notions. In other words, the diagnostic problem is synonymous to decision making (or classification) in this paper.

The structure of the paper is as follows. This section is an overview of basic concepts and facts. Section 3 describes the organisation of the tests. In Section 4, we give a short characteristics of 10 tested algorithms. Description of the data file from the field of rheumatology and experimental results in form of many tables are given in Section 5. Section 6 contains evaluation of the results and finally, Section 7 is the conclusion.

2.1. Basic setting

Let us suppose $(\Omega, \mathcal{X}, P)$ is a probabilistic space on which random variables η, ξ_1, ξ_2 $\cdots \xi_n$ are defined. *Diagnostic variable* η takes its values in a finite set of diagnoses $\{d_j\} = \boldsymbol{R}(\eta)$. (Symbol $\boldsymbol{R}(\vartheta)$ applied on a variable ϑ will denote its range (or codomain) in the sequel.) It is assumed the aim of the decision making is finding the most probable value of the η. All other variables, taking their values from finite sets denoted as $\boldsymbol{R}(\xi_1)$, $\boldsymbol{R}(\xi_2)\cdots$, $\boldsymbol{R}(\xi_n)$ are called *symptom variables* since their known values represent symptoms from which the unknown final diagnosis is inferred during decision making. Then, the set of all possible combinations of values of variables $\eta, \xi_1, \xi_2, \cdots \xi_n$ (i. e. their sample space), denoted as $\boldsymbol{R}(\eta, \xi_1, \xi_2, \cdots \xi_n)$, is a cartesian product of respective codomains:

$$\boldsymbol{R}(\eta,\, \xi_1, \xi_2, \cdots \xi_n) = \boldsymbol{R}(\eta) \times \boldsymbol{R}(\xi_1) \times \boldsymbol{R}(\xi_2) \cdots \boldsymbol{R}(\xi_n)$$

2.2. Idealized diagnostic problem

The mutual "behaviour" of η, $\xi_1, \xi_2, \cdots \xi_n$ is described completely by the joint distribution $P_{\eta\xi_1\xi_2\cdots\xi_n}$ induced from P and defined on $\boldsymbol{R}(\eta, \xi_1, \xi_2, \cdots \xi_n)$.

Suppose we are given the distribution $P_{\eta\xi_1,\xi_2\cdots\xi_n}$ and a subset $a = \{\xi_{i_1}, \xi_{i_2}, \cdots \xi_{i_k}\}$ of the set $\{\xi_1, \xi_2, \cdots \xi_n\}$ of all *symptom variables.* (Subset a is called *aperture* to stress it is a kind of filtering window through which we can see values of some symptom variables only during the decision making.) Then, the diagnostic problem can be formulated in the following way:

Diagnostic problem. Find the diagnosis $d_a(s_{i_1}, s_{i_2} \cdots s_{i_k})$ that is the most probable (according to the $P_{\eta\xi_1,\xi_2\cdots\xi_n}$) on the set

$$\{\omega \in \Omega \,|\, \xi_{i_1}(\omega) = s_{i_1} \,\&\, \xi_{i_2}(\omega) = s_{i_2} \,\&\, \cdots \xi_{i_k}(\omega) = s_{i_k}\} \tag{1}$$

for a given (i. e. observed) arbitrary combination $(s_{i_1}, s_{i_2} \cdots s_{i_k})$ of values of *symptom variables* from the set a.

A sequence of three steps providing an "obvious" solution to the *Diagnostic problem* may be denoted as

Algorithm $\mathbf{A}_0$:

Step 1: Marginalization of $P_{\eta\xi_1\xi_2\cdots\xi_n}$ to $P_{\eta\ \xi_{i_1}\xi_{i_2}\cdots\xi_{i_k}} = P_{\eta\xi_1\xi_2\cdots\xi_n}^{\eta\xi_{i_1},\xi_{i_2},\cdots\xi_{i_k}}$.

Step 2: Calculation of $|\eta|$ numbers representing the values of conditional probability $P_{\eta\,|\,\xi_{i_1}\xi_{i_2}\cdots\xi_{i_k}}(d_j\,|s_{i_1}, s_{i_2}\cdots s_{i_k})$ for individual diagnosis d_j and for the given combination $(s_{i_1}, s_{i_2}\cdots s_{i_k})$.

Step 3: Finding the optimal diagnosis $d_a(s_{i_1}, s_{i_2}\cdots s_{i_k})$:

$$d_a(s_{i_1}, s_{i_2}\cdots s_{i_k}) = \underset{d\,\in\,\boldsymbol{R}(\eta)}{\operatorname{argmax}}\ P_{\eta\,|\xi_{i_1}\xi_{i_2}\ldots\xi_{i_k}}(d|s_{i_1}, s_{i_2}\cdots s_{i_k}). \tag{2}$$

2.3. Approximations of the joint distribution

Leaving aside computational aspects, the $\boldsymbol{A}_0$ (as well as the presented diagnostic problem formulation) has one substantial drawback: We are never given the theoretical distribution $P_{\eta\xi_1\xi_2\cdots\xi_n}$ in full and directly. (The up to now discussion was just to expose the basic ideas and introduce notation.)

Therefore, the diagnostic problem has to be refined to cope with the "real world". To compensate for the loss of direct knowledge of $P_{\eta\xi_1\xi_2\cdots\xi_n}$, we expect to have some indirect information about $P_{\eta\xi_1\xi_2\cdots\xi_n}$ that will be called *knowledge base* and denoted by $\mathcal{K}$. It is done by postulating a set of conditions that we believe the theoretical $P_{\eta\xi_1\xi_2\cdots\xi_n}$ fulfills.

Instead of the unknown $P_{\eta\xi_1\xi_2\cdots\xi_n}$, we try to construct its approximation $\hat{P}_{\eta\xi_1\xi_2\cdots\xi_n}$ that could play its role in the diagnostic problem. The set of conditions $\mathcal{K}$ can define as feasible not only one distribution $\hat{P}_{\eta\xi_1\xi_2\cdots\xi_n}$, but a whole family $\mathcal{P}(\mathcal{K})$ of distributions complying with $\mathcal{K}$. There are many ways to define the set of postulated requirements representing the *knowledge base* $\mathcal{K}$. One of them and the only one used in this paper, is applying the concept of *marginal problem*.

2.4. Marginal problem formulation

Knowledge base $\mathcal{K}$ is given as a set of "small-dimensional" distributions (i. e. number of variables in the distribution is small; e. g. not exceeding 10.) postulated as theoretical *marginal distributions* of the $P_{\eta\xi_1,\xi_2\ldots\xi_n}$. This formulation of looking for $P_{\eta\xi_1,\xi_2\ldots\xi_n}$ is called *marginal problem*, see [7]. Here, the small-dimensional distributions are either explicitly given or calculated from *data D*. Instead of "small-dimensional distributions in $\mathcal{K}$", the one word term "oligodistributions" will be used in the sequel. This reflects the fact that they have usually a few of variables $s_{i_1}, s_{i_2}\cdots s_{i_k}$ and their respective sample spaces like $\boldsymbol{R}(\xi_l)$ consist of a few values only. (If the variables or sample spaces were not limited, though finite, there would be complexity problems with algorithms.) The second reason why the term *oligodistributions* is preferred to term *marginals* lies in the fact that small-dimensional distributions that are given by an expert as input need not be consistent and then, there does not exist any joint distribution whose marginals they might be.

Though it would be quite acceptable to select any arbitrary distribution from $\mathcal{P}(\mathcal{K})$, it is a common practice to use, beside $\mathcal{K}$, a general principle (e. g. maximal entropy principle, see [1, 5]) as an additional condition to force out uniqueness. Then, there exist algorithms (see e. g. see [1, 2] or [9]) that construct an approximation that is called *maximal entropy extension* (of a set of marginals). It should be stressed (see [1, 5]) that *maximal entropy extension*, beside being consistent with input oligodistributions, uses minimal additional information.

However, for the purpose of this paper, we take up the position that any numerical procedure that selects a distribution from $\mathcal{P}(\mathcal{K})$ or, at least, provides its conditional probability of type $P_{\eta\,|\,\xi_{i_1}\xi_{i_2}\cdots\xi_{i_k}}(d_j\,|s_{i_1}, s_{i_2}\cdots s_{i_k})$, is a decision making algorithm (denoted as) A_i. Their list is in the Section 4.

2.5. Input information for decision making

To summarize, we can encounter several types of objects when testing decision making algorithms.

1. *Knowledge base* $\mathcal{K}$ consists of a set of oligodistributions.

$$\mathcal{K} = \{o_1, o_2, \cdots o_l\}$$

   Individual oligodistributions are supposed to be provided by experts or generated from a statistical material, that is referred to as *data file* It should be stressed, though in general the notion of oligodistribution does not require it, that each oligodistribution in the knowledge base $\mathcal{K}$ *must* contain the *diagnostic variable* η and the fact is actively made use of in the algorithms. Just to explain, we presume that there is always more information about η in the distributions that contain it than in the distributions that describe mutual behaviour of symptom variables only and therefore, we do not let in such oligodistributions in the *knowledge base* $\mathcal{K}$ from the very start. As η is thus present by definition in all oligodistributions in the $\mathcal{K}$, it is superfluous to mention it explicitly and we will give only the names of symptom variables present in the oligodistributions in Tables 2 to 5.

2. Data or *data file* $\mathcal{D}$ is a set of combinations $(d, s_1, s_2, \cdots s_n,)$ of values of all random variables $\eta\,\xi_1\xi_2\cdots\xi_n$ that were measured or observed for a group of respondents r_i in the past. *Data file* $\mathcal{D}$ can be used, as mentioned above, for constructing oligodistributions in the *knowledge base* $\mathcal{K}$, for testing or for both purposes. (In general, there should be two separate *data file*s: One for building the oligodistributions, the other one for testing. In this paper, we consider only one *data file* $\mathcal{D}$ for both purposes!!)

3. *aperture* a is a subset $a = \{\xi_{i_1}, \xi_{i_2}, \cdots \xi_{i_k}\}$ of the set $\{\xi_1, \xi_2, \cdots \xi_n\}$ of all *symptom variables*. Only the values of symptom variables from the aperture are visible during the decision making.

4. *Facts* are values s_i of symptom variables from the aperture. In other words, they are fixed findings observed or measured on an individual respondent r for which we perform the act of decision making (i. e. we want to infer the value of the most probable diagnosis).

5. *Enforced sequence* (of oligodistributions) is an ordered subset of oligodistributions from *knowledge base* $\mathcal{K}$. For the testing purposes, we may wish not to use all the oligodistributions from $\mathcal{K}$. We may select only some of them and pretend that for a given testing run the *knowledge base* "shrunk" to the oligodistributions that are in the *enforced sequence*. The ordering on the subset is required because some algorithms are dependent on the way the oligodistributions are submitted to them and produce different results.
6. *Situation* s_i is an ordered pair $((o_{i_1}, o_{i_2}, \cdots o_{i_k}), a)$ (i.e. (enforced sequence, aperture)). This concept is useful as it describes in a unique way the situation under which the algorithms perform decision making. Definitions of the *situations* $s_1, s_2, \cdots s_{26}, s_{100}, s_{101}, \cdots s_{105}$ are in Tables 6 – 11. Then, the *situations* like s_i are used in Tables 12 – 16 to denote the rows in the tables.

3. TESTING SCHEME

To guarantee the same starting position for all tested algorithms, each of them has as its input the same *knowledge base* $\mathcal{K}$. Further, a *situation* s_i is defined by fixing an *enforced sequence* and an *aperture* a. This way, it is guaranteed that only certain oligodistributions from $\mathcal{K}$ are available. Then, a testing run for an algorithm A_j is performed in the following way:

For each object/respondent r_k in *data file* $\mathcal{D}$ the values of symptom variables from *aperture* a are submitted to the algorithm A_j and A_j performs the decision making that results in declaring an value $d(s_i, A_j)(r_k)$ from $\boldsymbol{R}(\eta)$ as the result of the decision making for the respondent r_k. If $d(s_i, A_j)(r_k)$ differs from $\eta(r_k)$ (as we know it from the *data file* $\mathcal{D}$), the counter of erroneous decisions (misclassifications) is augmented by one. After processing all respondents r_k from $\mathcal{D}$, the counter contains the total number of misclassifications $x_{(A_j, s_i, \mathcal{D})}$ that were committed by algorithm A_j on *data file* $\mathcal{D}$ under *situation* s_i. The same testing run is performed for another algorithm under similar conditions and results are presented in the Tables 12 – 16.

The objection that it is strange to use the same *data file* $\mathcal{D}$ both for generating oligodistributions and for testing can be rejected, as we are comparing different algorithms one against the other. Besides, one can hardly believe a procedure that would not do well under such conditions would be better when applied on other testing *data file* to which it were not adapted.

It should be stressed that this approach (i.e. using *situations*), describes the behaviour of algorithms in more details than if the comparison would take place just on all oligodistributions and with no filtering (by a).

4. COMPARED ALGORITHMS

There are 10 algorithms subjected to testing in this paper. Though the primal interest is to study algorithms suggested by A. Perez, to provide a contrast comparison another 3 programs were added. Namely, empirical distribution (whose misclassifications are denoted by symbol x_E in the Tables 12 – 16) and algorithms A_1, A_4 from

[8]. A. Perez suggested algorithms Exe (denoted as A_5) , $\overline{\text{Exe}}$ (denoted as A_6), DSS algorithm (A_7), asteroid algorithm (denoted as A_8) and a modification of A_6 that is presented in three versions (denoted as A_9, A_{10}, A_{11}). Some of these algorithms can be parameterized by parameter k and it is the case of the versions A_9, A_{10}, A_{11}.

1. x_E algorithm for evaluation of aposteriori conditional probability derived from *empirical distribution* $P_{E(\mathcal{D})}$ calculated from the *data file* $\mathcal{D}$. Number of erroneous decisions is denoted by x_E. No other algorithm can be better (i. e. with smaller number of errors) when applied to $\mathcal{D}$.
2. A_1 (described in [8]) is rather primitive and it is here for comparative purposes only. It is a normalized geometrical mean of aposteriori probabilities derived from all active oligodistributions. By definition, A_1 is not sensitive to changes in ordering of active oligodistributions.
3. A_4 (described in [8]) is doing quite well in various *situations*. It should not be too sensitive to changes in ordering of active oligodistributions (i. e. relatively independent of the fact whether RIP (Running intersection Property) condition is fulfilled or not.)
4. A_5 (described in [4]) is the Perez's algorithm Exe (Explicit Expression). In principle, it is evaluating a fraction where numerator consists of product of all non-empty "intersections" generated by odd number of oligodistributions. Similarly, the denominator of the fraction is a product of all non-empty "intersections" generated by even number of oligodistributions. The resulting vector (one number for each diagnosis) is normalized to 1.0, interpreted as aposteriori probability and used for decision making. Parameter k is used to limit the degree of "intersections". Thus, $k = 3$ means that only all basic active oligodistributions are used in the fraction as well as submarginals resulting from non empty intersections of all pairs and triplets of basic oligodistributions. With increasing number of oligodistributions, the parameter k should be increased as well. Not all the intersecting "suboligodistributions" are non empty, but it is obvious that k should be selected with grain of salt. E. g. for 11 oligodistributions and $k = 6$, there are 462 potential combinations contributing to the denominator in the fraction just for "intersections" with 6 oligodistributions. Though the size of intersecting "suboligodistributions" decreases, time and space demands go up with increasing k. In the Tables 12 – 16 in column with A_5 , the algorithm Exe has $k = 3$.
5. A_6 (described in [4]) is the Perez's algorithm $\overline{\text{Exe}}$ (i. e. normalized explicit expression). It is similar to A_5, but more complicated and leading to limitation on number of variables. It is necessary to calculate a vector of normalizing constants (one value for each diagnosis). The space for storing some results for all configurations of values of active symptom variables is required. Thus, the number of active variables should not exceed 20. The constants are used for multiplication of numbers from fraction formula before decision making. In the Tables 12 – 16 in column with A_6, the algorithm $\overline{\text{Exe}}$ has $k = 8$.
6. A_7 (described in [3] and partially in [4]) is the Perez's algorithm DSS (dependence structure simplification). It is reported to be dependent on the RIP condition fulfillment.

7. A_8 (described in [4]) is the Perez's *asteroid algorithm* that requires the input oligodistributions to have one part of variables in common. (These are called the *core*). The rest of variables in each oligodistribution is not present in other oligodistributions. With such star shaped structure (i. e. *asteroid*), the calculation is very fast.

8. A_9 is an algorithm by Perez similar to A_6, but the normalization constants are not calculated (being bottle-neck of the whole procedure). Apriori probability of the *diagnostic variable* η is used instead. This version of A_6 with "apriori multiplicants" is used with parameter $k = 1$.

9. A_{10} is the same algorithm as A_9, but with parameter $k = 2$.

10. A_{11} is the same algorithm as A_9, but with parameter $k = 3$.

5. EXPERIMENTAL RESULTS

All algorithms were tested on the data from the field of rheumatology (Prof. Rejholec, 1980). The *data file* $\mathcal{D}$ consists of 1089 patients and diagnosis variable η takes 4 different diagnoses. The file contains beside η, other 34 symptom variables ξ_i. Only 11 of them were used in the oligodistributions in the knowledge base $\mathcal{K}$, see Table 1. The symptom variables are of type gender, age, weight, working conditions etc. and their sample spaces have cardinality from 2 to 9.

The knowledge base $\mathcal{K}$ consists of 11 four-dimensional oligodistributions basically, see Table 2 and Table 3.

Two-dimensional oligodistributions $o_{12}, o_{13}, \cdots o_{16}$ (see Table 4), out of possible 45, were generated (derived) from oligodistributions $o_1, o_2, \cdots o_{11}$. Additional 7 four-dimensional oligodistributions $o_{101} - o_{107}$ are defined in Table 5.

Then, Tables $6, 7, 8, 9, 10, 11$ describe *situations* and Tables $12-16$ describe absolute number of errors x_{A_k} the algorithms achieve for the *situation* s_i. (Respective columns are denoted only by symbols A_k in the tables.)

Notation: If o_i is an oligodistribution, $\underline{o_i}$ is the set of symptom variables whose relation with the diagnosis variable η is described by o_i. The symbol $|\underline{o_i}|$ is the number of such symptom variables. Finally, $|o_i|$ stands for number of all combinations of symptom variables (or, in other words, number of atoms of set algebra created by symptom variables) in o_i. (Let us remind that due to postulated existence of η in all oligodistributions in $\mathcal{K}$, if the *space* e. g. for o_2 is 81, then the oligodistribution o_2 is given by 81×4 nonnegative numbers.) *Enforced sequences* are in the column entitled "oligodistributions" in Tables $6-11$.

Beside the number of misclassifications, each *situation* s_i is characterized by four additional variables. The symbol x_E denotes number of misclassifications committed by *empirical distribution* $P_{E(\mathcal{D})}$ derived from *data file* $\mathcal{D}$. The notion *active space* is the set of *symptom variables* that are present at least in one of the oligodistributions defined in *enforced sequence* and at the same time present in the aperture a. In other words, *active variables* lie in $a \cap \bigcup_i \underline{o_i}$.

Column *size* stands for the cardinality of value combinations from the *active space*. It is product of cardinalities of individual active *symptom variables*.

Symbol *nzAt* stands for non zero atoms in sample space of *symptom variables* from the *active space*. It should be always less or equal to the respective *size*.

Table 1. 11 symptom variables ξ_i active in $\mathcal{K}$.

sympt. variable	ξ_2	ξ_5	ξ_6	ξ_9	ξ_{14}	ξ_{15}	ξ_{20}	ξ_{21}	ξ_{22}	ξ_{23}	ξ_{31}
size	9	5	6	7	3	7	3	3	3	5	4

Table 2. Four-dimensional oligodistributions $o_1 - o_5$ in $\mathcal{K}$.

Oligodistribution \ variables	$\lvert\underline{o_i}\rvert$	ξ_{i_1}	ξ_{i_2}	ξ_{i_3}	ξ_{i_4}	space
o_1	4	ξ_2	ξ_5	ξ_9	ξ_{31}	1260
o_2	4	ξ_{14}	ξ_{20}	ξ_{21}	ξ_{22}	81
o_3	4	ξ_2	ξ_6	ξ_{15}	ξ_{23}	1890
o_4	4	ξ_2	ξ_{21}	ξ_{23}	ξ_{31}	540
o_5	4	ξ_6	ξ_9	ξ_{15}	ξ_{23}	1470

Table 3. Four-dimensional oligodistributions $o_6 - o_{11}$ in $\mathcal{K}$.

Oligodistribution \ variables	$\lvert\underline{o_i}\rvert$	ξ_{i_1}	ξ_{i_2}	ξ_{i_3}	ξ_{i_4}	space
o_6	4	ξ_9	ξ_{14}	ξ_{20}	ξ_{23}	315
o_7	4	ξ_5	ξ_6	ξ_9	ξ_{15}	1470
o_8	4	ξ_2	ξ_{14}	ξ_{20}	ξ_{31}	324
o_9	4	ξ_5	ξ_9	ξ_{14}	ξ_{15}	735
o_{10}	4	ξ_9	ξ_{15}	ξ_{22}	ξ_{31}	588
o_{11}	4	ξ_2	ξ_6	ξ_{14}	ξ_{21}	486

Table 4. Derived two-dimensional oligodistributions $o_{12} - o_{16}$ in $\mathcal{K}$

Oligodistribution \ variables	$\lvert\underline{o_i}\rvert$	ξ_{i_1}	ξ_{i_2}	space
o_{12}	2	ξ_2	ξ_5	45
o_{13}	2	ξ_2	ξ_9	63
o_{14}	2	ξ_5	ξ_9	35
o_{15}	2	ξ_2	ξ_{31}	28
o_{16}	2	ξ_5	ξ_{31}	20

Table 5. Four-dimensional oligodistributions $o_{101} - o_{107}$.

Oligodistribution \ variables	$\|o_i\|$	ξ_{i_1}	ξ_{i_2}	ξ_{i_3}	ξ_{i_4}	space
o_{101}	4	ξ_2	ξ_9	ξ_{15}	ξ_{20}	1323
o_{102}	4	ξ_2	ξ_9	ξ_{15}	ξ_{23}	2205
o_{103}	4	ξ_2	ξ_9	ξ_{15}	ξ_{31}	1763
o_{104}	4	ξ_{15}	ξ_{20}	ξ_{23}	ξ_{31}	420
o_{105}	4	ξ_9	ξ_{20}	ξ_{23}	ξ_{31}	420
o_{106}	4	ξ_2	ξ_{20}	ξ_{23}	ξ_{31}	540
o_{107}	4	ξ_2	ξ_9	ξ_{23}	ξ_{31}	1260

Table 6. Situations $s_1 - s_5$.

situation s_i	oligodistributions	aperture	active space	size
s_1	o_{12}	$\xi_1 \cdots \xi_{33}$	ξ_2, ξ_5	45
s_2	o_{12}, o_{13}	$\xi_1 \cdots \xi_{33}$	ξ_2, ξ_5, ξ_9	315
s_3	o_{12}, o_{13}, o_{14}	$\xi_1 \cdots \xi_{33}$	ξ_2, ξ_5, ξ_9	315
s_4	$o_{12}, o_{13}, o_{14}, o_{15}$	$\xi_1 \cdots \xi_{33}$	$\xi_2, \xi_5, \xi_9, \xi_{31}$	1260
s_5	$o_{12}, o_{13}, o_{14}, o_{15}, o_{16}$	$\xi_1 \cdots \xi_{33}$	$\xi_2, \xi_5, \xi_9, \xi_{31}$	1260

Table 7. Situations $s_6 - s_{11}$.

situation s_i	oligodistributions	aperture	active space	size
s_6	o_{13}	$\xi_1 \cdots \xi_{33}$	ξ_2, ξ_9	63
s_7	o_{14}	$\xi_1 \cdots \xi_{33}$	$\xi_5, \xi_9,$	35
s_8	o_{15}	$\xi_1 \cdots \xi_{33}$	ξ_2, ξ_{31}	36
s_9	o_{16}	$\xi_1 \cdots \xi_{33}$	ξ_5, ξ_{31}	20
s_{10}	o_{13}, o_{16}	$\xi_1 \cdots \xi_{33}$	$\xi_2, \xi_5, \xi_9, \xi_{31}$	1260
s_{11}	o_{13}, o_{15}	$\xi_1 \cdots \xi_{33}$	ξ_2, ξ_9, ξ_{31}	252

Table 8. Situations $s_{12} - s_{15}$.

situation s_i	oligodistributions	aperture	active space	size
s_{12}	o_{12}, o_{14}	$\xi_1 \cdots \xi_{33}$	ξ_2, ξ_5, ξ_9	315
s_{13}	o_1	$\xi_1 \cdots \xi_{33}$	$\xi_2, \xi_5, \xi_9, \xi_{31}$	1260
s_{14}	o_2	$\xi_1 \cdots \xi_{33}$	$\xi_{14}, \xi_{20}, \xi_{21}, \xi_{22}$	81
s_{15}	o_{13}, o_{14}	$\xi_1 \cdots \xi_{33}$	ξ_2, ξ_5, ξ_9	315

Table 9. Situations $s_{16} - s_{20}$.

sit. s_i	oligodistributions	aperture	active space	size
s_{16}	o_{18}	$\xi_1 \cdots \xi_{33}$	ξ_{14}, ξ_{20}	9
s_{17}	o_{23}	$\xi_1 \cdots \xi_{33}$	ξ_{21}, ξ_{22}	9
s_{18}	o_{18}, o_{23}	$\xi_1 \cdots \xi_{33}$	$\xi_{14}, \xi_{20}, \xi_{21}, \xi_{22}$	81
s_{19}	$o_3, o_4, o_6, o_7, o_5, o_1, o_{11}$	$\xi_1 \cdots \xi_{33}$	$\xi_2, \xi_5, \xi_6, \xi_9, \xi_{14}, \xi_{15}, \xi_{20}, \xi_{21}, \xi_{23}, \xi_{31}$	7144200
s_{20}	$o_3, o_4, o_6, o_7, o_5, o_1, o_2$	$\xi_1 \cdots \xi_{33}$	$\xi_2, \xi_5, \xi_6, \xi_9, \xi_{14}, \xi_{15}, \xi_{20}, \xi_{21}, \xi_{22}, \xi_{23}, \xi_{31}$	21432600

Table 10. Situations $s_{21} - s_{26}$.

situation s_i	oligodistributions	aperture	active space	size
s_{21}	$o_1, o_2, o_3, o_4, o_5, o_6, o_7$	$\xi_3 - \xi_8, \cdots \xi_{10} - \xi_{33}$	$\xi_5, \xi_6, \xi_{14}, \xi_{15}, \xi_{20}, \xi_{21}, \xi_{22}, \xi_{23}, \xi_{31}$	340200
s_{22}	$o_1, o_2, o_3, o_4, o_5, o_6, o_7$	$\xi_1 \cdots \xi_{33}$	$\xi_2, \xi_5, \xi_6, \xi_9, \xi_{14}, \xi_{15}, \xi_{20}, \xi_{21}, \xi_{22}, \xi_{23}, \xi_{31}$	21432600
s_{23}	o_1, o_2	$\xi_1 \cdots \xi_{33}$	$\xi_2, \xi_5, \xi_9, \xi_{14}$ $\xi_{20}, \xi_{21}, \xi_{22}, \xi_{31}$	102600
s_{24}	o_1	$\xi_1 \cdots \xi_{33}$	$\xi_2, \xi_5, \xi_9, \xi_{31}$	1260
s_{25}	o_1, o_2, o_3	$\xi_1 \cdots \xi_{33}$	$\xi_2, \xi_5, \xi_6, \xi_9, \xi_{14}, \xi_{15}, \xi_{20}, \xi_{21}, \xi_{22}, \xi_{23}, \xi_{31}$	21432600
s_{26}	o_1, o_2, o_3, o_5, o_6	$\xi_1 \cdots \xi_{33}$	$\xi_2, \xi_5, \xi_6, \xi_9, \xi_{14}, \xi_{15}, \xi_{20}, \xi_{21}, \xi_{22}, \xi_{23}, \xi_{31}$	21432600

Table 11. Situations $s_{100} - s_{105}$.

situation s_i	oligodistributions	aperture	active space	size
s_{100}	$o_1 - o_{11}$	$\xi_2, \xi_9, \xi_{15}, \xi_{20}, \xi_{23}, \xi_{31}$	$\xi_2, \xi_9, \xi_{15}, \xi_{20}, \xi_{23}, \xi_{31}$	26460
s_{101}	$o_{101} - o_{107}$	$\xi_2, \xi_9, \xi_{15}, \xi_{20}, \xi_{23}, \xi_{31}$	$\xi_2, \xi_9, \xi_{15}, \xi_{20}, \xi_{23}, \xi_{31}$	26460
s_{102}	$o_{101}, o_{102}, o_{103}$	$\xi_2, \xi_9, \xi_{15}, \xi_{20}, \xi_{23}, \xi_{31}$	$\xi_2, \xi_9, \xi_{15}, \xi_{20}, \xi_{23}, \xi_{31}$	26460
s_{103}	o_{101}, o_{107}	$\xi_2, \xi_9, \xi_{15}, \xi_{20}, \xi_{23}, \xi_{31}$	$\xi_2, \xi_9, \xi_{15}, \xi_{20}, \xi_{23}, \xi_{31}$	26460
s_{104}	o_{101}	$\xi_2, \xi_9, \xi_{15}, \xi_{20}, \xi_{23}, \xi_{31}$	$\xi_2, \xi_9, \xi_{15}, \xi_{20}$	1323
s_{105}	o_{107}	$\xi_2, \xi_9, \xi_{15}, \xi_{20}, \xi_{23}, \xi_{31}$	$\xi_2, \xi_9, \xi_{23}, \xi_{31}$	1260

Table 12. Behaviour of $A_1, A_4, A_5 - A_{11}$ for situations $s_1 - s_7$.

situation	x_E	size	nzAt	A_1	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}
s_1	515	45	41	515	515	515	515	515	515*	535	535	535
s_2	441	315	205	505	491	491	491	505	491*	508	505	505
s_3	441	315	205	504	485	561	485	504	559	497	485	485
s_4	385	1260	324	489	450	542	452	489	539	492	672	452
s_5	385	1260	324	487	454	739	451	487	670	490	818	542
s_6	516	63	62	516	516	516	516	516	516*	529	529	529
s_7	612	35	30	612	612	612	612	612	612*	630	630	630

Table 13. Behaviour of $A_1, A_4, A_5 - A_{11}$ for situations $s_8 - s_{14}$.

situation	x_E	size	nzAt	A_1	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}
s_8	515	36	26	515	515	515	515	515	515	534	534	534
s_9	618	20	15	618	618	618	618	618	618*	625	625	625
s_{10}	385	1260	324	518	518	518	498	518	498	522	522	522
s_{11}	471	252	132	502	486	486	486	502	486*	504	504	504
s_{12}	441	315	205	524	514	514	514	524	514	530	514	514
s_{13}	385	1260	324	385	385	385	385	385	385*	387	387	387
s_{14}	596	81	36	596	596	596	596	596	596*	607	607	607

Table 14. Behaviour of $A_1, A_4, A_5 - A_{11}$ for situations $s_{15} - s_{20}$.

situation	x_E	size	nzAt	A_1	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}
s_{15}	441	315	205	506	496	496	497	506	496*	526	510	510
s_{16}	632	9	7	632	632	632	632	632	632*	637	637	637
s_{17}	652	9	7	652	652	652	652	652	652*	655	655	655
s_{18}	596	81	36	622	622	622	607	622	607*	648	648	648
s_{19}	24	7E6	1046	208	162	159		208	186	211	214	165
s_{20}	17	21E6	1057	216	170	146		216	194	218	230	147

Table 15. Behaviour of $A_1, A_4, A_5 - A_{11}$ for situations $s_{21} - s_{26}$.

situation	x_E	size	nzAt	A_1	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}
s_{21}	172	340200	773	481	457	512	445	481	481*	493	793	441
s_{22}	17	21432600	1057	216	172	146		216	189	218	230	147
s_{23}	114	102060	864	363	363	363	363	363	363	381	381	381
s_{24}	385	1260	324	385	385	385	385	385	385*	387	387	387
s_{25}	17	21432600	1057	275	242	242		275	265	275	256	256
s_{26}	17	21432600	1057	247	201	209		247	214	253	248	187

Table 16. Behaviour of $A_1, A_4, A_5 - A_{11}$ for situations $s_{100} - s_{105}$.

situation	x_E	space	nzAt	A_1	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}	A_{11}
s_{100}	230	26460	640	429	387	469	369	428	441	446		464
s_{101}	230	26460	640	290	262	281	246	290	285	289	378	326
s_{102}	230	26460	640	296	265	265	265	296	265*	300	329	281
s_{103}	230	26460	640	315	292	292	290	315	292	315	302	302
s_{104}	350	1323	362	350	350	350	350	350	350	354	354	354
s_{105}	422	1260	234	422	422	422	422	422	422	434	434	434

6. EVALUATION OF EXPERIMENTAL RESULTS

We tried to characterize the experimental results by verifying validity of some assertions representing some trends in the tables. Unfortunately, almost none of the assertions holds in a "logical form" i. e. without exceptions. But, one may expect that some assertions may hold reasonably often to justify using some deductions or strategies. First, some trends common for all algorithm are presented and then, comparation of "discernment power" of individual algorithms.

6.1. Trends resulting from situations

1. Each algorithm A decides better for situations with smaller empirical error x_E.
$$\bigvee_{A \in \mathcal{A}} \bigvee_{s_u, s_v} [(x_E(s_u) \leq x_E(s_v)) \implies (x_A(s_u) \leq x_A(s_v))]$$
This assertion is not true. See e. g. s_{19} and s_{20}. It looks like the composition of oligodistributions can be more important then the smaller number of empirical errors x_E. However, the implication holds for quite a lot of situations.

2. Increasing an oligodistribution results in better decision for knowledge bases consisting of one oligodistribution only.

$$\bigvee_{A\in\mathcal{A}} \bigvee_{\mathcal{K}:|\mathcal{K}|=1} \bigvee_{o_1\in\mathcal{K}} \bigvee_{o_2:\underline{o_2}\subset\underline{o_1}} \bigvee_{s} [x_A^{o_1}(s) \leq x_A^{o_2}(s)]$$

   It seems to be valid. E.g. $\underline{o_1}(s_{13}) \supset \underline{o_{12}}(s_1), \underline{o_{14}}(s_7), \underline{o_{15}}(s_8), \underline{o_{16}}(s_9)$ and $x_A^{o_1}(s_{13}) = 385 < x_A^{o_{12}}(s_1) = 515, x_A^{o_{14}}(s_7) = 612, x_A^{o_{15}}(s_8) = 515, x_A^{o_{16}}(s_9) = 618$ for both $A = A_1$ and $A = A_4$.

3. Larger size of space of active variables results in smaller errors.

   Let $Ac(\mathcal{K})$ be a set of active variables and $|Sp(\mathcal{K})|$ be the size of the space $Sp(\mathcal{K})$ of atoms of the set algebra created by active symptom variables.

$$Ac(\mathcal{K}) = \bigcup_{o_i\in\mathcal{K}} \underline{o_i} \cap a, \quad |Sp(\mathcal{K})| = \prod_{\xi_i\in Ac(\mathcal{K})} |\xi_i|$$

   then

$$\bigvee_{A\in\mathcal{A}} \bigvee_{\mathcal{K}_1,\mathcal{K}_2} Sp(\mathcal{K}_1) \leq Sp(\mathcal{K}_2) \Longrightarrow x_A(\mathcal{K}_2) \leq x_A(\mathcal{K}_1)$$

   This assertion is not true. See e.g. s_{19} vs. s_{20} and s_{14} vs. s_8. Namely, $Sp(s_{19}) = 7144200 < Sp(s_{20}) = 21432600$ and $x_{A_1}(s_{19}) = 208 < x_{A_1}(s_{20}) = 216$ and $x_{A_4}(s_{19}) = 162 < x_{A_4}(s_{20}) = 170$. Similarly, $Sp(s_{14}) = 81 > Sp(s_8) = 36$ and $x_A(s_8) = 515 < x_A(s_{14}) = 596$ for both $A = A_1$ and $A = A_4$. On the other hand, it holds very often. $Sp(s_{15}) = 315 > Sp(s_{18}) = 81 > Sp(s_{16}) = 9 \geq Sp(s_{17}) = 9$ and $x_{A_1}(s_{15}) = 506 < x_{A_1}(s_{18}) = 622 < x_{A_1}(s_{16}) = 632 < x_{A_1}(s_{17}) = 652$ and $x_{A_4}(s_{15}) = 496 < x_{A_4}(s_{18}) = 622 < x_{A_4}(s_{16}) = 632 < x_{A_4}(s_{17}) = 652$. However, having in mind the mentioned exceptions, it is obvious that a concrete composition of oligodistributions can be more important than the mere number of active symptom variables $Ac(\mathcal{K})$ and the size of the space $Sp(\mathcal{K})$.

6.2. Comparison of individual algorithms

Again, the assertions hold with certain probability. Some deductions result from the testing runs not presented in this paper.

1. In general, A_5, A_6, A_{11} obtain very good results (small number of misclassifications)

2. A_{10} is worse than others due to the non fitted parametrization i.e. $k = 2$. Usually, versions with odd k behave better.

3. A_4 provides certain robustness (independence on the situations) and, in general, good results.

4. A_1 is the worst one what could be expected due to his simplicity. On the other hand, its behaviour can be taken as a natural upper bound on misclassifications.

5. A_7 (DSS) and A_8 (asteroid) perform well if the underlying structure of oligodistributions fulfills the required conditions e. g. RIP-conditions or asteroid structure. There are in certain disadvantage when compared on the general structures with algorithms like ExE, Ex2 and A4, that are not so sensitive in this respect.

6. A_6 seems to improve with more complex structure till the moment some internal limits are trespassed and the program abends. In such situations, an appropriate version of the Ex3 seems to be good replacement. But, all algorithms cloned from ExE should have parametrization of k fitted to the number of oligodistributions in the used *enforced sequence*.

7. Differences between the worst and best algorithm can achieve up to 25 %.

8. No algorithm outperforms the others (in Pareto sense) in all *situations*.

9. Narrowing the applied aperture a can have a devastating effect on the oligodistributions in *enforced sequence* which can contain the same reduced oligodistributions. Then, active oligodistributions not only shrink but their number decreases, too. The result is that misclassifications for all algorithms are alike.

7. CONCLUSIONS

1. Three new algorithms A_8 (Asteroid), A_5 (ExE), A_6 ($\overline{\text{ExE}}$) and A_9, A_{10}, A_{11} (clones of Ex3) were suggested by Albert Perez. Beside having good theoretical foundations, they outperform the up to now best heuristical algorithm A_4 for various *situations*. A_8 is optimal for special structure of oligodistributions in the $\mathcal{K}$.

2. A_5 is a universal very efficient algorithm and A_6 is recommendable for sizes up to 1000000 atoms where it yields the smallest error among all tested algorithms.

3. The idea to test the algorithms using the concept of *situations* seems to be justified as it gives a more complex view of their behaviour in practice.

4. The algorithms described in [4] use relative entropy $H(P, \hat{P})$ (and terms like *multiinformation* and *multiinformation content*) to find the "closest" approximation $\hat{P}_{\eta\xi_1\xi_2\cdots\xi_n}$. The testing scheme uses as a natural measure of efficiency the number of erroneous decisions (misclassifications). The fact that the algorithms are so successful means that there is a high correlation between both approaches.

5. The future activity should be aimed on procedures for selecting the fitting k parameter.

6. Other aim could be a procedure for selecting the most discerning individual oligodistributions and selecting the groups of oligodistributions with the greatest "synergy".

7. The procedures are to be looked for that would suggest changing the structure of the given oligodistributions to a new one fulfilling special requirements (RIP-conditions, asteroid structure) with minimal number of changes like adding/removing a variable to/from an oligodistribution, throwing out oligodistributions from the active *enforced sequence*, changing the ordering etc.

ACKNOWLEDGEMENTS

The author is indebted to Milan Studený for providing preliminary version of [4].

(Received August 24, 2006.)

REFERENCES

[1] P. Cheeseman: A method of computing generalized Bayesian probability values of expert systems with probabilistic background. In: Proc. 6th Joint Conf. on AI(IJCAI-83), Karlsruhe.

[2] W. E. Deming and F. F Stephan: On a least square adjustment of sampled frequency table when expected marginal totals are known. Ann. Math. Stat. *11* (1940), 427–444.

[3] A. Perez: ε-admissible simplifications of the dependence structure of random variables. Kybernetika *13* (1979), 439–449.

[4] A. Perez and M. Studený: Comparison of two methods for approximation of probability distributions with prescribed marginals. Kybernetika *43* (2007), 5, 591–618

[5] E. T. Jaynes: On the rationale of maximum-entropy methods. Proc. IEEE *70* (1980), 939–952.

[6] R. Jiroušek, A. Perez, and O. Kříž: Intensional way of knowledge integration for expert systems. In: DIS'88 – Distributed Intelligence Systems, Varna 1988, pp. 219–227.

[7] H. G. Kellerer: Verteilungsfunktionen mit gegebenen Marginalverteilungen. Z. Wahrsch. verw. Gebiete *3* (1964), 247–270.

[8] O. Kříž: A new algorithm for decision making with probabilistic background In: Trans. Eleventh Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, Vol. B, Prague 1990, Academia, Prague 1992, pp. 135–143.

[9] O. Kříž: Optimizations on finite-dimensional distributions with fixed marginals. In: WUPES 94, Proc. Third Workshop on Uncertainty Processing (R. Jiroušek, ed.), Třešť 1994, pp. 143–156.

Otakar Kříž, Renoirova 620, 152 00 Praha 5. Czech Republic.
e-mail: kriz_v@volny.cz, kriz@ceps.cz