POSSIBLE-WORLDS SEMANTICS FOR RULE-BASED EXPERT SYSTEMS WITH SET-VALUED WEIGHTS

IVAN KRAMOSIL

The contribution deals with expert systems the deduction machines of which are based on production rules with observations as antecedents and hypotheses as succedents. To each rule or sentence in the databasis its lower and upper weights or degrees of validity are ascribed, the values of these weights being formulas of an appropriate propositional language interpreted as sets of worlds in which the sentences are certainly (lower weight) or possibly (upper weight) valid. A simple way is presented how to deduce the weights of a given hypothesis, and how to decide about its acceptability, given some empirical data (observations, e.g.).

1. INTRODUCTION

For the sake of this short contribution we may take *expert system* as a software collection consisting of *data*, *rules* and *inference engine* which enables to deduce new statements from the given data by an appropriate use of the rules. Data are sentences of a language describing empirical facts being at our disposal (observations, results of experiments, ...). Only very simple rules will be taken into consideration, namely implication-like formulas $A \rightarrow C$, where A is a conjunction $A_1 \wedge A_2 \wedge \ldots \ldots \wedge A_n$ of empirical facts and C is a disjunction $C_1 \vee C_2 \vee \ldots \vee C_m$ of particular elementary formulas called *hypotheses*.

At this level the inference machine is nothing else than a purely logical deduction or its fragment working within given time and space limitations. However, what is typical of expert systems is the fact that neither the rules nor the data can be stated as certain, there is always a degree of uncertainty or weight of validity connected with each rule and datum. One of the main philosophical and methodological problems of expert systems is how to quantify these uncertainties or weights.

The difficulties connected with the most common numerical weights, usually taking their values within a finite interval of reals, are well known and it is not our aim to discuss them here in details. They arise mainly from the problem how to combine the weights ascribed to data and rules used during a deduction in order to obtain the weight with which the resulting statement holds. The extensional approach combines the weights according to a fixed combination procedure so that the resulting weight depends on the data and rules used during the deduction just through their weights. This viewpoint usually results in relatively simple computations and in simple explanations how the values have been obtained (supposing the user asks for such an explanation). On the other hand, it reduces very strictly our possibilities to interpret weights as probabilities with which data and rules hold, as this approach is not able to reflect all the richess and flexibility of mutual statistical dependences among data and rules. The intensional approach, at least in its probabilistic version, treats all the weights as marginal or conditional probabilities uniquely defined by and computable from the simultaneous probability distribution covering all the data and hypotheses in question as one vector-valued random variable. In case the assumption of existence of the simultaneous distribution is ontologically justified, this viewpoint seems to be acceptable, but it leads to computational difficulties if the simultaneous distribution in question and the corresponding marginal and conditional probabilities are to be computed or estimated. As suggested by Bundy [1], the contradictory positions of the two approaches need not be taken as absolute but rather as resulting from the numerical nature of the weight functions. In other words said, ascribing to data and rules weights with values in a richer and more flexible space than the unit interval, we would be able to express statistical dependences by weight values themselves, so preserving the extensional nature of the combination procedure computing the output weights. Probably the most intuitive and straightforward idea is to consider weights as taking their values in the space of all subsets of a fixed nonempty set – universe of discourse. The idea was investigated in [2], cf. also [3], at a purely set-theoretical level; in what follows we shall discuss it from a more semantical point of view.

2. SOME ALTERNATIVES FOR POSSIBLE-WORLDS SEMANTICS

In the Kolmogorov probability theory probabilities are numerical values of functions defined on certain subsets (measurable sets or random events) of a fixed abstract nonempty space Ω. Hence, each numerical weight ascribed to a datum or rule and supposed to be interpretable as probability in the Kolmogorov sense, must be a value taken by a probability measure on a measurable subset of Ω. So the idea arises to define the value of the weight in question just by the corresponding subset of Ω. Obviously, the advantages of both the extensional as well as the intensional approaches are preserved. Or, ascribing subsets $e(A)$ and $e(B)$ of Ω to formulas (data or rules) A and B as their weights, the weights $e(A) \cup e(B)$ for $A \vee B$ (disjunction), and $e(A) \cap e(B)$ for $A \wedge B$ (conjunction) are defined extensionally through the weights $e(A)$, $e(B)$, but reflect the statistical relations between A and B. E.g., $e(A \cup B) = e(A) \cup e(B) = e(A)$, if $e(A) \supset e(B)$, i.e., if $B \rightarrow A$ holds, on the other hand, $e(A) \cup e(B)$ is "maximum", if A and B are mutually exclusive, i.e. if $A \rightarrow \neg B$ and $B \rightarrow \neg A$ hold. Like as in the axiomatic probability theory, the abstract nature of the set Ω enables to transform probabilities of qualitatively different events onto a common unifying level so that they may be easily combined with each other.

However, these advantages may quickly become disadvantages supposing the designer or user of an expert system is to ascribe set weights to data and rules in question. Or, not having at hand any semantical interpretation of the points in Ω, i.e., not being able to ascribe any extra-mathematical meaning to them, the user must know a priori all the statistical dependences between the corresponding events to be able to ascribe the set values appropriately. E.g., considering the example in [1] with $\Omega = \{0, 1, \ldots, 99\}$ and two formulas A, B with probabilities $p(A) = 0{\cdot}75$, $p(B) = 0{\cdot}25$, the expert, having ascribed $\{0, 1, \ldots, 74\}$ as weight to A, must know the statistical dependence between A and B to decide, whether to ascribe either $\{0, 1, \ldots, 24\}$, or $\{0, 4, 8, \ldots, 92, 96\}$, or even $\{75, 76, \ldots, 99\}$, or something else to B. Hence, from the viewpoint of practical applications the situation looks like a vicious circle the only escape from which is to consider the points in Ω as endowed by meaning. Ascribing subsets of Ω to formulas, the expert takes substantial profit of their semantics not considering explicitly their statistical dependences, he or she does so rather implicitly, just through the meanings. In [3] we consider an ecological example of dependences between occurrences of various species of plants in a territory. Here weights ascribed to corresponding implications are simply subsets of this territory where the expert believes (or has observed) the dependence in question to hold. These weights can be drafted into a (copy of) map of the territory and combined with other drawings, e.g., by a superprojection, the statistical dependences between various assertions being implicitly hidden in the drawings.

In this example, as well as in a number of other ones, the space Ω can be identified with an extra-mathematical universe of actual topological places, each of them being called a "possible world". However, the idea of possible worlds goes much further and a "possible world" can be defined as the subset of a universe of discourse, where some sentences $S_1, S_2, \ldots, S_n$ of an appropriate language take given truth values, so that we have 2^n-element space of possible worlds supposing the sentences are logically independent. To take subsets of such a space of possible worlds as weights ascribed to data and rules is the same as to take, in the role of weights, the propositional formulas built from the sentences $S_1, \ldots, S_n$ and defining the sets of worlds in question. Hence, weights can be understood as more conditions expressed in, maybe, different language than that in which data and rules are, and conditioning, in the expert's opinion, the validity of the data and rules in question. In the rest of this contribution this interpretation of rule-based expert systems with set-valued weights will be described in more details.

3. PROPOSITIONAL SENTENCES AS SET-VALUED WEIGHTS

Consider three sequences $\mathcal{O} = \langle o_1, o_2, \ldots \rangle$, $\mathcal{H} = \langle h_1, h_2, \ldots \rangle$, and $\mathcal{W} = \langle w_1, w_2, \ldots \rangle$ of propositional indeterminates (no assumptions concerning their mutual disjointness or inclusion being taken in this moment). Let $\mathcal{L}_1$ ($\mathcal{L}_2$, $\mathcal{L}_3$, resp.) be the propositional language generated by the indeterminates in $\mathcal{O}$ ($\mathcal{H}$, $\mathcal{W}$, resp.) and by the propositional connectives $\neg$ (negation), $\vee$ (disjunction), $\wedge$ (conjunction), $\rightarrow$ (implication) and $\equiv$ (equivalence) in the usual way, let $\mathcal{L}_4$ be generated by $\mathcal{O} \cup \mathcal{H} \cup \mathcal{W}$ in the same way. By L_i, $i = 1, 2, 3, 4$, denote the set of all well-formed formulas of the language $\mathcal{L}_i$, so that $L_i \subset L_4$ for $i \leqq 3$. *Data* are formulas from L_1, *rules* are formulas of the form $A \rightarrow C$ with $A \in L_1$ and $C \in L_2$, *weights* are formulas from L_3. The most simple and easy to interprete are data of the form $o_{i_1} \wedge o_{i_2} \wedge \ldots$ $\ldots \wedge o_{i_n}$ and rules of the form $(o_{i_1} \wedge i_{i_2} \wedge \ldots \wedge o_{i_n}) \to (h_{j_1} \vee h_{j_2} \vee \ldots\ldots \vee h_{j_m})$. Finally, consider a fixed *model* or *relation structure* $\mathfrak{M}$ such that all indeterminates from $\mathcal{O} \cup \mathcal{H} \cup \mathcal{W}$ are ascribed just one truth value by $\mathfrak{M}$, hence, for each formula $\phi \in L_4$ either $\mathfrak{M} \models \phi$ or $\mathfrak{M} \models \neg\phi$ (either ϕ or its negation are valid, i.e. ascribed the truth value TRUE, in $\mathfrak{M}$ or by $\mathfrak{M}$). By T_i (F_i, resp.), $i \leqq 3$, we shall denote a fixed proposional tautology (negation of T_i, resp.) in L_i, evidently, T_1, T_2, T_3 are also tautologies and F_1, F_2, F_3 negations of tautologies in L_4.

Taking profit of the Dempster-Shafer idea of two-values weights, cf. [2], we ascribe to each formula E from the database (datum or rule) a pair $\langle W_*(E), W^*(E)\rangle \in$ $\in L_3 \times L_3$ of formulas such that $\mathfrak{M} \models (W_*(E) \to E) \wedge (E \to W^*(E))$; $W_*(E)$ is the *lower weight* ascribed to F and $W^*(E)$ is the *upper weight*. More generally, a formula may occur twice or more times in the database with different weights (obtained, say, from different experts), but in every case $W_*(E)$ is *objectively* (i.e. w.r. to $\mathfrak{M}$) a sufficient and $W^*(E)$ a necessary condition for E. Hence, the expert's subjectivity is expressed by his choice of these two conditions and his uncertainty is expressed by the fact that $W^*(E) \to W_*(E)$ need not hold, in $\mathfrak{M}$, except of ideal or optimal cases (the inverse implication being valid trivially). Evidently, if $\langle E, W_*(E),$ $W^*(E)\rangle$ is in the database, $\neg E$ can be joined with this database together with $W_*(\neg E) = \neg W^*(E)$ and $W^*(\neg E) = \neg W_*(E)$, moreover, each formula E can be joined with database together with trivial weights $W_*(E) = F_1$ and $W^*(E) = T_1$.

Let us write $E_1 \leqq E_2$, if $\mathfrak{M} \models (E_1 \to E_2)$ for two formulas $E_1, E_2 \in L_4$. Evidently, if $\langle E, W_{*1}(E), W_1^*(E)\rangle$ and $\langle E, W_{*2}(E), W_2^*(E)\rangle$ are two occurrences of a formula E in the database such that $W_{*1}(E) \leqq W_{*2}(E)$ and $W_2^*(E) \leqq W_1^*(E)$, then $\langle E, W_{*1}(E), W_1^*(E)$ can be erased from the database without any loss of information. More generally, for each two occurrences $\langle E, W_{*1}(E), W_1^*(E)\rangle$ and $\langle E, W_{*2}, W_2^*(E)\rangle$ we can easily deduce that $\mathfrak{M} \models (W_{*1}(E) \vee W_{*2}(E)) \to E$ and $\mathfrak{M} \models E \to (W_1^*(E) \wedge W_2^*(E))$, so that $\langle E, W_{*1}(E) \vee W_{*2}(E), W_1^*(E) \wedge W_2^*(E)\rangle$ can be add to the database (and the former two occurrences can be erased due to the argumentation above). A more powerful system of such rules can be obtained by an easy re-formulation of the ten rules And 1–6 and Not 1–4, from [1], from the set-theoretical languages into the propositional one ($\tilde{W}_*(E)$ and $\tilde{W}^*(E)$ denote the improved lower and upper weights).

(R 1) $$\tilde{W}^*(E) = (\neg W_*(\neg E)) \wedge W^*(E)$$

(R 2) $$\tilde{W}_*(E) = (\neg W^*(\neg E)) \vee W_*(E)$$

(R 3) $$\tilde{W}^*(\neg E) = (\neg W_*(E)) \wedge W^*(\neg E)$$

(R 4) $$\tilde{W}_*(\neg E) = (\neg W^*(E)) \vee W_*(\neg E)$$

(R 5) $$\tilde{W}^*(E) = [W^*(E \wedge F) \vee (\neg W_*(F))] \wedge W^*(E)$$

(R 6) $$\tilde{W}_*(E) = W_*(E \wedge F) \vee W_*(E)$$

(R 7) $$\tilde{W}^*(F) = [W^*(E \wedge F) \vee (\neg W_*(E))] \wedge W^*(F)$$

$$(\text{R } 8) \qquad \tilde{W}_*(F) = W_*(E \wedge F) \vee W_*(F)$$

$$(\text{R } 9) \qquad \tilde{W}^*(E \wedge F) = \tilde{W}^*(E) \wedge W^*(F) \wedge W^*(E \wedge F)$$

$$(\text{R } 10) \qquad \tilde{W}_*(E \wedge F) = [W_*(E) \wedge W_*(F)] \vee W_*(E \wedge F)$$

If necessary or convenient, the rules for other propositional connectives can be easily deduced from (R 1)–(R 10). In each case $\tilde{W}^*(A) = W^*(A) \wedge \ldots$ and $\tilde{W}_*(A) = W_*(A) \vee \ldots$, so that $\tilde{W}^*(A) \leqq W^*(A)$ and $\tilde{W}_*(A) \geqq W_*(A)$. When applying these rules we may take profit of the algorithm presented in [1], but we shall not go into details. What is of importance is the fact that no information concerning the data, rules, and their mutual dependences is lost during the transformations by (R 1)–(R 10), in spite of the inevitable loss of information through each extensional manipulation with numerical weights.

The presented model offers a degree of freedom to experts, users or knowledge engineers when deciding which conditions, connected with the validity of a hypothesis, will be explicitly stated as premises in a rule and which will play the role of weights ascribed to this rule. Or, consider a rule $A \to C$ with weights $\langle W_*, W^* \rangle$ and suppose that $W_*, W^* \in L_1$ (hence, the expert conditions the validity of C by the validity of A together with some specified results of certain supplementary tests, treatments or observations, on the other hand he knows that some other results, namely those covered by $\neg W^*$, of maybe different tests, treatments or observations make the validity of C impossible). So $\mathfrak{M} \models W_* \to (A \to C)$, hence, $\mathfrak{M} \models (W_* \wedge A) \to C$, which implies $\mathfrak{M} \models T_3 \to [W_* \wedge A) \to C]$ and $\mathfrak{M} \models [(W_* \wedge A) \to C] \to T_3$, recall that $T_3 \in L_3$ is a tautology. Hence, a new rule $W_* \wedge A \to C$ can be joined with the database with the trivially optimal weights $\langle T_3, T_3 \rangle$, so that this rule is surely valid w.r. to $\mathfrak{M}$. On the other hand, $\mathfrak{M} \models (A \to C) \to W^*$, so that $\mathfrak{M} \models C \to W^*$ as well; combining this result with $\mathfrak{M} \models (W_* \wedge A) \to C$ already proved we obtain, that the hypothesis C itself can be joined with our database together with weights $\langle W_* \wedge A, W^* \rangle$. Evidently, for each $\langle E, W_*(E), W^*(E) \rangle$ in the database, $\mathfrak{M} \models W_*(E) \to W^*(E)$, hence, $\mathfrak{M} \models \neg W_*(E) \vee W^*(E)$. This idea of a free trade-off between conditions and weights stands close to that of "inversely-driven" expert systems presented in [4], when the system outputs, which premises and with which (numerical) weights should be verified so that the given conclusions (hypotheses) could be stated with desired weights.

On the other hand, the language $\mathscr{L}_3$ can be richer than $\mathscr{L}_1$ so that also the "true" randomness concerning the validity of data or rules can be expressed in our frameworks. This is reached in case the language $\mathscr{L}_3$ contains sentences expressing results of random experiments or values taken by random variables. In principal, this is always possible supposing the experiments or variables take values in a finite set, then the abstract space on which they are formally defined can be also taken as finite and can be embedded into the set of possible worlds generated by an appropriate language $\mathscr{L}_3$. In a sense, an event can be *defined* as random with respect to $\mathscr{L}_1$ supposing the set of worlds when this event occurs cannot be expressed as a set of worlds definable by $\mathscr{L}_1$, hence, neither the complete knowledge of truth values taken by all sentences from $\mathscr{L}_1$ suffices to decide, whether the random event in question has actually occurred or not.

In expert systems with numerical weights a hypothesis is, as a rule considered as acceptable supposing we are able to ascribe to it a weight sufficiently great, most often, sufficiently close to one or to other maximum value. Within our framework we may consider two formulas, say U_*, $U^* \in L_3$, playing the role of threshold values in the sense that a hypothesis C with values $W_*(C)$, $W^*(C)$ is accepted iff $\mathfrak{M} \models (U_* \to W_*(C)) \wedge (U^* \to W^*(C))$. Formula U_* may be interpreted, in terms of heuristics, as defining the set of "typical" worlds for which validity of C should be assured, formulas $\neg U^*$ defines, under the same interpretation, the set of "non-typical" or even "pathological" cases for which the hypothesis C may be known not to be valid without affecting its acceptability. Even if C is accepted according to such a criterion, its application in an actual world may not be free of an uncertainty or risk. It is caused by the fact that actual worlds can be classified only on the ground of the language $\mathscr{L}_1$, so that we may not be sure whether U_* and U^* actually hold (the same situation as when using heuristics proved for typical or non-pathological cases in a real world).

4. CONCLUSIONS

The aim of the model presented above is twofold. First, to enable, for experts, to express their knowledge, opinions and beliefs in a non-numerical, say, verbal form and, second, to handle with the obtained information at this non-numerical level as far as possible, postponing the definitions of numerical weights to data and rules, if necessary, till the moment when this can be done with the minimum risk or under minimum extra-conditions. (E.g., a condition A may occur in some rules concerning a hypothesis C, but it may disappear, due to an appropriate combination of rules and data, from the final weights ascribed to C so that it is not necessary to ascribe numerical weights to A and $\neg A$ as requested, when starting our manipulations immediately at a numerical level. Because of the limited extent of this contribution and with the aim to emphasize its discussion-open nature we have intently kept the text on a rather philosophical level not going into a deeper formalization and technical details.

REFERENCES

[1] A. Bundy: Incidence calculus — a mechanism for probabilistic reasoning. J. Automat. Reasoning *1* (1985), 3, 263—283.

[2] A. P. Dempster: A generalization of Bayesian inference. J. Roy. Statist. Soc. Ser. B *30*(1968), 208—241.

[3] I. Kramosil: Expert systems with non-numerical belief functions. Problems Control Inform. Theory *17* (1988), 4, 285—295.
[4] J. Rakušan and H. Šafránková: Expert system with a non-traditional form of inference (in Czech). In: Aplikace umělé inteligence AI '88 — sborník referátů, Praha, 1988, pp. 109—116.

RNDr. Ivan Kramosil, DrSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.