MODELING PARTIAL IGNORANCE IN ARTIFICAL INTELLIGENCE APPLICATIONS
DALĒJĀS NEZINĀŠANAS MODELĒŠANA MĀKSLĪGA INTELEKTA PIELIETOJUMOS

OLEG UZHGA-REBROV

Rezekne Higher Education Institution
Atbrivoshanas al. 90, Rezekne LV – 4600, Latvia,
Phone: + 371 9472205, *e-mail*: rebrovs@tvnet.lv

Abstract. *This study aims to extend and deepen a survey of modern extensions of probability theory represented in [6, 7]. The classical probability theory possesses rather limited possibilities and cannot cope with situations of partial ignorance. Other approaches are required allowing one to solve tasks of that kind. The given paper considers the generalised approach to modelling partial ignorance and its interpretation in the terms of upper and lower probabilities, second order probabilities and belief functions.*

Keywords: *belief functions, focusing, partial ignorance, revision, second order probabilities, upper and lower probabilities.*

Introduction

Artificial intelligence (AI) is commonly treated as different methods of knowledge representation and processing. The present study considers knowledge both as information obtained from objective sources (measurements, observations, and database) and subjective evaluations, judgements and experts' findings. The size and quality of actual knowledge in the context of the specific task solution may vary from complete knowledge to total ignorance [7]. A situation when some knowledge about the subject area is available but it is not complete and/or reliable will be called a situation of partial ignorance. In order to model and process partial knowledge successfully, it is necessary to have a clear idea about the sources and reasons of partial and total ignorance. Those sources and reasons can be divided into three large classes [5].

1. Incompleteness of knowledge represents situations when values of certain variables of the task are unknown. For example, when a database of a group of individuals is being created it might be known that a certain individual is a student, whereas his name might be unknown. In this situation the value of variable Individual's Name is unknown.
2. Imprecision of knowledge relates to the case when values of all or some variables cannot be determined with the required extent of precision. In the above mentioned task of database creation, what is only known regarding the value of variable Age for a specific individual is that he is over 30 years.
3. Uncertainty of knowledge covers the situations when propositions are not certainly true. For the case of database construction, a proposition Individual X is a student - made on the basis of the evidence of the person not sure about that - can serve as an example of uncertain proposition.
4. The concepts of knowledge incompleteness, imprecision and uncertainty have to be strictly distinguished from the concept of uncertainty (randomness) of various processes of the actual world. The external uncertainty exists objectively and does not depend on human consciousness. Uncertainty in AI represents the incompleteness and uncertainty of our knowledge (information) about real world events and processes. For example, assigning chances of random event occurrence might turn to be a troublesome problem for an individual though he/she might clearly imagine the character and structure of those events. Situations are however possible when an individual has no clear ideas about the nature and structure of random events or processes. The uncertainty of that kind - related to the absence of complete and reliable knowledge is examined in this paper.

ISBN 9984 – 779 – 06 – 8

For a long time, probability theory served as the only means for managing uncertainties of the actual world. Probability theory proves to be a very effective mathematical apparatus, however it can be used in the presence of these two conditions:

5. A complete space of random events is specified.
6. To each event a unique number is attached that evaluates its chances and is called probability.

As soon as one of those conditions is broken, probability theory stops being an adequate mathematical apparatus for dealing with uncertainty [3, 6]. Other - more suitable approaches to modelling knowledge incompleteness and uncertainty are becoming necessary. Some of approaches of that kind are discussed in the present paper.

General requirements to the approaches to modelling partial ignorance

In case if the available knowledge about the real world subject area is not reliable and/or certain, our first task is to model such knowledge. Before considering specific approaches to solving that task, let us formulate general requirements that have to be satisfied by any of those approaches. It is evident that the main requirement is an adequate representation of the actual state of things in a form that could help quantify our extent of ignorance. (Qualitative approaches to solving that task exist, for example, nonmonotonic logics, default logics, theory of potential surprise, etc. However, they are not considered in the present paper). In [1, 2] the requirements are represented in the most developed and justified form. Let there be a set of propositions Q about the actual states of the subject area of the actual world. Consider a subset $R(g) \subseteq Q$ of that set. If complete and true knowledge is available, the subset R(g) can be easily divided into two subsets : $\tau(g) = \{gi = T\}$, $\theta(g) = \{gi = F\}$, where symbols T and F denote the truth or falsity of the respective proposition. If our knowledge is not complete, we can distinguish two subsets of propositions $\tau^*(g) \in R(g)$ and $\theta^*(g) \in R(g)$, regarding whose status of truth or falsity we are certainly sure, and a subset of propositions $V(g) = R(g) \setminus (\tau^*(g) \cap \theta^*(g))$, about the truth/falsity status of which we are not certainly sure. Then a subset $\tau^*(g) \cup V(g)$ will contain the propositions which are certainly true (a subset $\tau^*(g)$), and propositions $gi \in V(g)$, which are possibly true. Similarly, a set $\theta^*(g) \cup V(g)$ will contain the propositions which are certainly false (a subset $\theta^*(g)$) and propositions $gi \in V(g)$ which are possibly false.

How to model such a state of partial ignorance? Assume that we are able to construct a function ζ, which attaches the number equal to 1 to all elements from of $\tau^*(g)$; the number equal to 0 to all elements of $\theta^*(g)$ and some number $\alpha \in [0,1]$ to all elements from V(g). Whatever function ζ is, it has to meet the following requirements:

– the value $\zeta(gi \vee gj)$ must be determined by values $\zeta(gi)$ and $\zeta(gj)$ only;
– the value $\zeta(gi \wedge gj)$ must be determined by values $\zeta(gi)$ and $\zeta(gj)$ only;
– the value $\zeta(\neg gi)$ must be determined by $\zeta(gi)$ only.

If some probabilistic function stands for function ζ, $\zeta(gi \vee gj) = \zeta(gi) + \zeta(gj)$, if and only if gi, gj $\in \tau^*(g)$ and gi, gj are probability independent. In general case, function $\zeta(gi \vee gj)$ is not a function of $\zeta(gi)$, $\zeta(gj)$. Similarly, function $\zeta(gi \wedge gj)$ is not a function of $\zeta(gi)$, $\zeta(gj)$. The only requirement the probability function of that kind satisfies is $\zeta(\neg gi) = \zeta 1$ (gi). From this it follows that no probabilistic function whatever could model partial ignorance in general case. Two evaluation functions are needed for that. Suitable for that purpose are functions of this kind:

$$SN(g) = \begin{cases} 1, \text{ if } g \in \tau^*(g), \\ 0, \text{ otherwise.} \end{cases}$$

ISBN 9984 – 779 – 06 – 8

$$S\Pi(g) = \begin{cases} 1, \text{ if } g \in \tau^*(g) \cup V(g), \\ 0, \text{ otherwise.} \end{cases}$$

The interpretation of these functions is simple enough and is logically validated. SN(g) = 1 means that g is a certainly true proposition, whereas SΠ(g) = 1 means that g may potentially be a true proposition, though it is not excluded that it might be false. Such an interpretation of SΠ function represents our partial ignorance regarding the truth status of proposition g. Functions SN and SΠ are partially compositional with respect to operations of logical intersection, union and negation:

$SN(gi \wedge gj) = \min (SN(gi), SN(gj))$;

$S\Pi(gi \vee gj) = \max (S\Pi(gi), S\Pi(gj))$;

$SN(g) = 1 - S\Pi(\neg g)$.

Function SN, however, cannot be compositional regarding the disjunction of propositions and function SΠ is not compositionl with regard to the conjunction of propositions.

It is easy to check that $SN(g) = 1 \Rightarrow S\Pi(g) = 1$, and that $g \in V(g)$, if and only if SN(g) = 0 and SΠ(g) = 1. From this it follows that the truth status of proposition g is correctly modelled by a pair of figures:

(SN(g) = 1, SΠ(g) = 1) – g is certainly true;

(SN(g) = 0, SΠ(g) = 0) – g is certainly false;

(SN(g) = 0, SΠ(g) = 1) – the status of g is not known for sure (the state of partial ignorance).

In what follows we will discuss possibilities of function SN and SΠ interpretation in the context of existing approaches to modelling states of partial ignorance.

Interpretation of functions SN and SΠ

Upper and lower probabilities

Upper and lower probabilities are used to model those situations of partial ignorance when ignorance is related to the probabilistic structure. A complete knowledge about the space of random events is available, but due to the absence of reliable information it is only possible to assign a set of probability distributions P = {Pi} in this space. Any distribution $Pi \in P$ can be a true distribution and there is no reason to prefer some $Pi \in P$ with respect to all other distributions in P. The maximum entropy principle for selecting single distribution – recommended by some authors- is not a valid choice criterion in general case. In such a situation, functions SN and SΠ are naturally interpreted as follows:

$S\Pi(P) = \sup_{P \in P} P$:

$SN(P) = \inf_{P \in P} P$.

Despite the clear and unambiguous interpretation, operating upper and lower probabilities is fairly complicated. A more detailed information on possible techniques of managing probabilities of that kind can be found in [8].

Second order probabilities

Using upper and lower probabilities to model partial ignorance related to the probabilistic structure provides rather limited possibilities to the user. If possible, it would be more suitable to introduce a structure in the set of probabilistic evaluations, which would allow one to represent conclusions regarding the relative plausibility of particular evaluations. Such a structure might be introduced in different ways. One of the techniques is to treat a set of probabilistic evaluations P as a fuzzy set and form the membership function μ(P) in that set which would represent

ISBN 9984 – 779 – 06 – 8

subjective judgements of the individual about the plausibility of particular evaluations. Another way allowing not to exceed probability theory is to form and employ a second-order probability distribution. This distribution is formed in the set of probabilistic evaluations and represents uncertain judgements of the individual about plausibility extent of particular evaluations. The second-order probabilities are frequently called metaprobabilities. Metaprobabilities are not suitable for modelling uncertain judgements in the set of objective probability evaluations. A structure of that kind represents the objective state of things. The only proper tool for removing uncertainty is acquiring additional objective information. Using metaprobabilities proves to be appropriate in the cases when a set (interval) of possible probabilistic evaluations is obtained in a subjective way and the individual has some reasons to consider certain evaluations to be more plausible than others. It should be noted that the application of mathematical expectations of second-order probability distributions does not remove the prior uncertainty and only masks it (see, for example, [4]).

Belief functions

The theory of belief functions - also known as Dempster-Shafer theory - has been developed to model a specific kind of partial ignorance, with which probability theory cannot cope in principle. The conceptual statement of probability theory is represented by a condition that the space of random events is specified completely and uniquely, and that each event is assigned a probabilistic evaluation. In case if probabilities can only be assigned to certain subsets of events from the general space, probability theory cannot be employed in principle to modelling such a specific kind of partial ignorance.

According to the theory of belief functions, to each subset $X \subseteq \Omega$, belief mass m(X) can be assigned that represents the extent of the individual's support of truth X. Then the total amount of support in favour of the truth of subset $A \subseteq \Omega$ is expressed by a belief function as follows:

$$bel(A) = \sum \emptyset \neq X \subset A \; m(X).$$

The values of belief functions are interpreted as probabilities in the original Dempster-Shafer theory. Belief functions possess the following features:

$$bel(\emptyset) = 0;$$

$$bel(\Omega) = 1;$$

$$bel(A1 \cup A2 \cup \ldots \cup An) = \sum i \; bel(Ai) \; - \sum i > j \; bel(Ai \cap Aj) - \ldots - (-1)n \; bel(A1 \cap A2 \cap \ldots \cap An).$$

(Note, that the last expression also holds for the classical probabilities with the change of sign $\geq$ to strict equality $=$).

Plausibility extent of subset A, pl(A), quantifies the maximal quantity of potential support which could be given to A:

$$pl(A) = \sum X \cap A \neq \emptyset \; m(X) = bel(\Omega) - bel(\hat{A}).$$

Using the general approach to partial ignorance modelling discussed in the previous section, one can correctly state that function bel(A) is interpreted as function SN(A), whereas function pl(A) is interpreted as function SΠ(A).

The theory of belief functions is a widely recognised tool for managing a specific kind of partial ignorance. Ph. Smets has developed his own version of that theory assuming that no probability distribution exists in the space of random events Ω. He has simply examined beliefs at the so-called credal level. He has called his version of belief functions *transferable belief model*. The mathematical apparatus of the transferable belief model at the credal level considerably corresponds to that of Dempster-Shafer theory. The so-called pignistic transformation of beliefs represents the benefit of transferable belief models. As a result of that transformation, belief evaluations at the credal level are transformed into the so-called pignistic probabilities, which could be manipulated using general rules of probability theory. The pignistic probabilites can be successfully applied for decision analysis and choice under risk. It should be noted that the pignistic probabilities in no case can be interpreted as usual probabilities. They represent specific

ISBN 9984 – 779 – 06 – 8

evaluations obtained by artificial transformation of probabilities and show all signs of probabilistic evaluations. As a matter of fact they are not probablities in the commonly accepted sense.

Problems of probabilistic reasoning in situations of partial ignorance

In most common sense, knowledge (complete and incomplete) can be divided into two large groups [1]:

1) based on factual evidence; and
2) generic knowledge.

Factual evidence ensures knowledge about the state of the actual world in the specific situation. This evidence might have different nature and its truth degree can vary within very wide limits. Generic knowledge relates to some subject area as a whole and does not refer to particular situations. Generic knowledge can be represented in the form of a set of plausible rules, different relationships (functional, stochastic etc.), probability distributions and so on. To give an idea of different kinds of knowledge, let us consider the physician's knowledge. His generic knowledge consists in generalised links between the symptoms and the diseases and includes data on the disease frequency (distribution) in the population. Factual evidence for the physician consists in the the symptoms and results of laboratory investigation of the pacient.

Conditioning the existing knowledge on the basis of new data (evidences) is an important step for removing partial ignorance. Two principally different types of conditioning are represented by focusing and revision [1]. The essence of focusing is conditioning generic knowledge by the factual evidence. In the above example with symptoms and deseases, focusing might yield a change of the reference class of diseases, which became possible in the presence of symptoms observable at the pacient. In general probability theory, focusing is accomplished by using Bayes' theorem.

Unlike focusing, revision is either conditioning the generic knowledge by another piece of generic knowledge (G-revision) or conditioning the factual evidence by another piece of factual evidence (F-revision). In the above example with diseases and symptoms, the result of G-revision might consist in modifying the physician's knowledge about disease-symptom links by using new medical knowledge (say, publications in special journals or statistical data). F-revision might consist in the change of the physician's beliefs about a disease, say, by learning the results of laboratory investigation.

In probability theory, both G- revision and F-revision are implemented by means of Bayes' theorem. Thus, the same mathematical apparatus is employed both for focusing and revision in probability theory. The implementation of focusing on the basis of Bayes' theorem seems to be quite justified. A classical task of re-evaluation of probabilities of hypotheses can be considered as a standard focusing task. The recalculation of the posterior probabilities of the hypotheses in the presence of new evidence is successfully accomplished using Bayes' theorem. As regards revision tasks, the use of Bayes' theorem does not seem to be so evident. Let us consider general statement of Bayes' theorem.

$$P(A/B) = P(B \cap A)/P(A).$$

The advocates of standard Bayesian approach assert that as a result of using that rule a change of reference class A does not occur. They treat the posterior distribution P(B/A) as a result of reduction of reference class from the whole space of events up to subset A. That interpretation coincides with the definition of focusing, but not revision, as it is stated frequently.

Now let us discuss approaches to solving tasks of focusing and revision under partial ignorance. In belief function theory, revision is carried out on the basis of Dempster combination rule. That rule actually combines the conjunctive revision of data under partial ignorance with Bayesian normalisation rule [1]. The essence of Dempster rule of conditioning can be expressed as relating the available belief masses mi for focal elements $Ei \in \Omega$ to subsets $Ei \cap A$, where A is a new portion of information, with further re-normalisation of belief masses in those extended subsets.

ISBN 9984 – 779 – 06 – 8

The expected degree of potential support of B provided that A is true, can be calculated as
$S\Pi(B/A) = S\Pi(A \cap B) / S\Pi(A)$.
That rule can also be set in the general context of upper and lower probabilities as follows:

$$S\Pi(B/A) = \sup \{P(A \cap B) / P((A), P \leq S\Pi, P(A) = S\Pi(A)\} \quad (1)$$

The last expression represents the essence of revision under partial representation technique of that kind.
An alternative conditioning rule can be given in this form:

$$S\Pi A(B) = S\Pi(A \cap B) / (S\Pi(A \cap B) + SN(A \cap \bar{B})). \quad (2)$$

That rule represents focusing. The underlying idea is to calculate possibility of B under the assumption that A is true not making any propositions as to how the set of probabilities will be revised.
In [1] it is shown that function SΠA (B) is generally less informative than function SΠ (A / B) and can even be less informative than the initial function SΠ (A). This result seems to be a paradox. It is, however, easy to explain it. Let a set of probabilities P represent our partial knowledge about the actual value of probability P. Hence SΠA (B) is a degree of potential belief that B is an element (subset) of A, which can be treated as the modification of generic knowledge (focusing). On the other hand, the value of SΠ (A / B) is simply a modification of the initial information, which is a typical revision. SΠA (B) can broaden the interval of probabilities of SΠ (A). The process of focusing can be justified in terms of functions SΠA(B) only, and SΠA(B) can be viewed of as the upper limit of a family of belief functions derived by transferring the belief masses mi to subsets Ei ∩ A. The latter provides an explanation of the above result of function SΠA(B) informativity decrease. Due to that result, an important remark has to be made. Belief function theory has been developed in order to combine uncertain evidences. In other words, it was designed to implement F-revision only. Any attempts to apply that theory to solving focusing tasks lead to unsatisfactory results.

Conclusions

Active research and development of new approaches to modelling partial ignorance was seen lately. That is caused by practical needs. Currently it is widely accepted that classical probability theory possesses limited possibilities regarding modelling and processing incomplete and uncertain knowledge. To quantify the degree of ignorance, two values are required. In the most general case these can be represented as the values of necessity function SN and of possibility function SΠ. The interpretation of functions SN and SΠ depends on the specific approach to partial ignorance modelling. By upper and lower probabilities partial ignorance related to the probabilistic structure is correctly modelled. Belief and plausibility functions are successfully employed to model partial ignorance that refers to the structure of random event space. Using metaprobabilities is an attempt to structurise partial uncertain knowledge in the space of probability evaluations.
The classification of probabilistic reasoning tasks into tasks of focusing and revision helps correctly evaluate the suitability degree of each specific approach to modelling incomplete knowledge. In classical probability theory all kinds of probabilistic reasoning tasks are solved on the basis of Bayes' theorem. Its use for solving focusing tasks seems to be quite justified, whereas its application to revision tasks is not so evident.
Belief function theory faces serious difficulties when it is employed to solve focusing tasks. It is much better adapted to solving revision tasks as was planned by its founders.

ISBN 9984 – 779 – 06 – 8

References

1. D.Dubois, H.Prade, Ph.Smets. Representing Partial Ignorance http://iridia.ulb.ac.be/~psmets/Rep_Partial_ignorance
2. P.E.Lehner, K.B.Laskey, and D.Dubois. An Introduction to Issues in Higher Order Uncertainty. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, Vol.26, No 3, May 1996, pp. 289 – 293
3. A.Mosleh and V.M.Bier. Uncertainty about Probability: A Reconciliation with the Subjectivist Viewpoint. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, Vol.26, No 3, May 1996, pp. 303 – 310
4. R.E.Neapolitan. Is Higher-Order Uncertainty Needed? IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, Vol.26, No 3, May 1996, pp. 294 – 302
5. Ph.Smets. Varieties of ignorance and the need for well-founded theories. http://iridia.ulb.ac.be/~psmets/Variety Ignorance
2. O.Uzhga-Rebrov. Uncertain Probabilities. Proceedings of the 4th International Conference „Environment. Technology. Resourses”, June 26-28, Rezekne, Rezekne Higher Education Institution, pp. 377 - 384
3. O.Uzhga-Rebrov. A Survey of Modern Extensions of Probability Theory. Proceedings of the International Conference „Scientific Achievments for Wellbeing and Development of Society”, March 4-5, 2004, Rezekne, Rezekne Higher Education Institution, pp. 107 – 112
4. P.Walley. Statistical Reasoning with Imprecise Probabilities. Thomson Press (India) Ltd, New Delphi, 1991

ISBN 9984 – 779 – 06 – 8