REASONING WITH CONSISTENT PROBABILITIES

Dimiter Driankov

Department of Information and Computer Science
University of Linkoping
8-581 83 Link8ping, Sweden

ABSTRACT

The objective of the present paper is twofold: first, to establish explicit conditions for the elicitation of consistent *a priori* and *conditional* probabilities for a set of *events* representing pieces of *evidence* and *hypotheses.* Furthermore, an algorithm is proposed which uses these consistent input probabilities to compute *lower* and *upper bounds* for higher order *joint* probabilities. Secondly, problems concerning the *aggregation* and *propagation* of probabilistic estimates are considered. It is shown how these could be solved by using the higher order *joint* probabilities obtained for the elements of the so-called *complete set of mutually exclusive atomic events.*

I INTRODUCTION

A representation of uncertainty in terms of classical Bayesian probabilities requires three basic inputs, namely: (i) a set of n *events* $\{E_i\}$, each *event* designating either an *evidence*, if it is in the premise of an IF - THEN rule, or a *hypothesis*, if it appears as a conclusion in such a rule.; (ii) *a priori* probabilities $P(E_i)$, $1 \leq i \leq n$ for the *events* under consideration; (iii) *conditional* probabilities $P(E_i / E_j)$ and $P(E_i / \tilde{E}_j)$.

A number of problems are associated with this type of representation of uncertainty. First, the above input probabilities elicited from the expert may not be consistent with respect to the following two conditions (1) and (2) respectively:

$$P(E_i) = P(E_j).P(E_i / E_j) + (1 - P(E_j)).P(E_i / \tilde{E}_j)$$

$$P(E_i E_j) = P(E_i / E_j).P(E_j) = P(E_j / E_i).P(E_i)$$

Second, (1) and (2) are not fulfilled for arbitrary *conditional* probabilities; further requirments for consistency are needed reflecting the relationship between the $P(E_i)$'s and the $P(E_i / E_j)$'s. Third, if the input probabilities are not consistent, then the problem of *propagation* of uncertainty can not be modeled by the following classical rule:

$$P(E_i / E^*_j) = P(E_i / E_j).P(E_j / E^*_j) + P(E_i / \tilde{E}_j).P(\tilde{E}_j / E^*_j) \quad (3)$$

So, what are the consistency conditions under which we can apply (3) ? Finally, consider questions like: How to determine a consistent $P(E_k / E_i E_j)$, representing a rule as "IF E_i and E_j THEN E_k ; How to determine a consistent higher order *joint* probability such as $P(E_i E_j E_k)$ etc. These are usually answered under the very much restrictive assumption that all E. 's represent *independent evidence.* Also it is often forgotten that although the input probabilities might have been elicited in a consistent way, the resulting higher order *joint* probabilities might not be consistent at all.

II Obtaining consistent input probabilities

Once a set of n *events* $\{E_i\}$ is specified together with the *a priori* probabilities $\{P(E_i)\}$, one is asked to assess the *conditional* probabilities $P(E_i / E_j)$ and $P(E_i / \tilde{E}_j)$, thus obtaining the following two matrices: (i) the *dependent occurence matrix* DO = $[a_{ij}]$, $1 \leq i \leq n$, $1 \leq j \leq n$, $a_{ii} = 1$, with $a_{ij} = P(E_j / E_i)$;

(ii) the *dependent non-occurence matrix* DNO = $[b_{ij}]$, $1 \leq i \leq n$, $1 \leq j \leq n$, $b_{ii} = 1$, with $b_{ij} = P(E_j / \tilde{E}_i)$.

Now if E_i and E_j are two *independent evidences* then, $a_{ij} = P(E_j / E_i) = P(E_j)$; if they are two *mutually exclusive hypotheses*, then $a_{ij} = P(E_j / E_i) = 0$. Thus, knowledge about *independent evidence* and *mutually exclusive hypotheses* is explicitly encoded in the DO-matrix. The problem is that two subjective assessments, say, $P(E_j / E_i)$ and $P(E_i / E_j)$ might be inconsistent with respect to (2) while $P(\tilde{E}_i / E_j)$ and $P(E_i / \tilde{E}_j)$ might be inconsistent with respect to (1). An additional trouble, concerning implementation, is reflected in the huge number of subjective assessments: in the case of the DO-matrix their number is n.(n-1); n.(n-1) in the case of the DNO-matrix, plus of course the n $P(E_i)$'s. However, what seems to have been completely ignored is that all $P(E_i / E_j)$'s for $i > j$ in the DO-matrix, can be computed from the $P(E_i / E_j)$'s for $i < j$ by rewriting (2) as:

$$P(E_i / E_j) = (P(E_j / E_i).P(E_i))/P(E_j) \quad (4)$$

Unfortunately, applying (4) produces consistent $P(E_i / E_j)$'s only if introducing some further restrictions on $P(E_j/E_i)$ i.e. $P(E_j/E_i) \in [LB_{ji}, UB_{ji}]$ where

$$LB_{ji} = \max(0, (P(E_i) + P(E_j) - 1))/P(E_i))$$

$$UB_{ji} = \min(1, P(E_j)/P(E_i))$$

The question now is: If $P(E_j/E_i) \in [LB_{ji}, UB_{ji}]$ will $P(E_i/E_j)$ be also within the range of its admissible values ? The answer is, yes, iff $P(E_j/E_i) \in [LB_{ji}, UB_{ji}]$ and $P(E_i/E_j)$ is obtained by (4). Thus, given the $P(E_i)$'s and the $P(E_i/E_j)$'s for i<j, the latter within the range of their admissible values, all remaining $P(E_i/E_j)$'s for i>j can be obtained automatically from (4) and vice versa. Also, the number of the subjectively assessed conditional probabilities is reduced to n.(n-1)/2 . As to the DNO-matrix, rewriting (1) as,

$$P(E_i/\tilde{E}_j) = (P(E_i) - P(E_j).P(E_i/E_j))/P(\hat{E}_j) \quad (5)$$

we can compute consistent entries using the $P(E_i)$'s and the consistent entries of the DO-matrix, thus further n.(n-1) subjective assessments being dropped aside.

To conclude we will finally show that the consistent values obtained for $P(E_i)$, $P(E_j)$, $P(E_i/E_j)$ and $P(E_i/\tilde{E}_j)$ satisfy the condition under which (3) is fullfilled.

Proposition . Let $P(E_i/E_j)$ defines the point $(1, P(E_i/E_j))$ in the positive quadrant of a coordinate system, while $P(E_i)$ and $P(E_j)$ define the point $(P(E_j), P(E_i))$ and $P(E_i/\tilde{E}_j)$ defines a point $(0, P(E_i/\tilde{E}_j))$. Then, if $P(E_i/\tilde{E}_j)$ is obtained by (5) the point $(0, P(E_i/\tilde{E}_j))$ belongs to the straight line that passes through the points $(P(E_j), P(E_i))$ and $(1, P(E_i/E_j))$.

III Constructing joint probabilities

The entries of the DO and the DNO matrices can not be used for representing rules of the form, IF E_i and E_j THEN E_k . Thus, there is a need for determining a consistent higher order *conditional* probability such as $P(E_k/E_i E_j)$ which is not directly assessed by the expert, but can be computed as,

$$P(E_k/E_i E_j) = P(E_i E_j E_k)/P(E_i E_j) \quad (6)$$

Here, we have to determine a consistent $P(E_i E_j E_k)$, given the already consistent *a priori* and *conditional* probabilities related to E_i, E_j and E_k . This in turn, can be used to obtain bounds for $P(E_k/E_i E_j)$ by use of (6).

In (E. Kounias 1968) it has been proved that the *best upper* (UB) and *lower* (LB) bounds for a *union of events* i.e. $P(E_1 \cup E_2 \cup ... \cup E_n)$ are given as:

$$UB = \min\left(\sum_{i=1:n} P(E_i) - \max_{j=1:n}\left(\sum_{i \neq j, i=1:n} P(E_i E_j)\right), 1\right)$$

$$LB = \max\left(\sum_{i=1:n} P(E_i) - \sum_{i=1:n} \sum_{j=i+1:n} P(E_i E_j)\right)$$

However, we are interested in bounds on higher order *joint* probabilities while the above result is valid for bounds on *union of events*. However, this is not a serious obstacle since one has that:

$$P(\tilde{E}_1 \tilde{E}_2 ... \tilde{E}_n) = 1 - P(E_1 \cup E_2 \cup ... \cup E_n) \quad (11)$$

Hence, we wish to obtain bounds on $P(D_1 D_2 ... D_n)$ where, each D_i can be either E_i or $\tilde{E}_j$. To be able to achieve this, we require bounds on $P(\tilde{D}_1 \cup \tilde{D}_2 \cup ... \cup \tilde{D}_n)$ using as input the following already consistent probabilities: (i) the $P(E_i)$'s; (ii) the $P(E_i/E_j)$'s, and (iii) the $P(E_i/\tilde{E}_j)$'s. Then the algorithm proposed is as follows:

Step 1. Compute a matrix C as follows:

If $\tilde{D}_i = E_i$ and $\tilde{D}_j = E_j$: $c_{ij} = P(E_i E_j) = P(E_i/E_j).P(E_j)$

If $\tilde{D}_i = E_i$ and $\tilde{D}_j = \tilde{E}_j$: $c_{ij} = P(E_i \tilde{E}_j) = P(E_i/\tilde{E}_j).P(\tilde{E}_j)$

If $\tilde{D}_i = \tilde{E}_i$ and $\tilde{D}_j = E_j$: $c_{ij} = P(\tilde{E}_i E_j) = P(E_j/\tilde{E}_i).P(\tilde{E}_i)$

If $\tilde{D}_i = \tilde{E}_i$ and $\tilde{D}_j = \tilde{E}_j$: $c_{ij} = P(\hat{E}_i \hat{E}_j) = (1 - P(E_i/\tilde{E}_j)).P(\tilde{E}_j)$

Step 2 . The quantities $P(C_i)$ are computed as: $P(C_i) = P(\tilde{D}_i)$

Step 3 . The *upper bound* UB is to be obtained as:

a) Compute $\sum_{j \neq i, j=1:n} c_{ij}$ i =1, 2, ... ,n

b) Compute $\sum_{i=1:n} P(C_i)^*$

c) Compute the *upper bound* UB according to:

$$UB = \min\left(\sum_{i=1:n} P(C_i) - \max\left(\sum_{j \neq i, j=1:n} c_{ij}\right), 1\right)$$

Step 4 .

a) Compute a set of *lower bounds* SLB as follows:

SLB(1,i) = $P(C_i)$ i = 1,2,...,n

SLB(2,i,j) = $P(C_i) + P(C_j) - c_{ij}$ i, j>i = 1,2,...,n

$$SLB(3,i,j,k) = P(C_i) + P(C_j) + P(C_k) - c_{ij} - c_{ik} - c_{jk} \quad i, j>i, k>j = 1,2,\dots,n$$

...

$$SLB(n,i,j,k,\dots) = \sum_{i=1:n} P(C_i) - \sum_{i=1:n} \sum_{j=i+1:n} c_{ij}$$

b) Compute the *best lower bound* LB as the maximum of all SLB's

Finally, once we have obtained both LB and its corresponding UB we can use (11) and thus obtain:

$$1 - UB \leq P(D_1 D_2 \dots D_n) \leq 1 - LB$$

It is to be stressed here that the bounds on the higher order *joint* probabilities might be consistent for some of them and inconsistent for others. To resolve this problem will require to determine explicit conditions for the input probabilities so, that the bounds on all higher order *joint* probabilities be consistent. This in turn will produce consistent higher order *conditional* probabilities. The consistency condition for each particular higher order *joint* probability is that, $LB \leq UB$ i.e. an inequality which two parts are linear functions of certain particular *a priori* probabilities and certain entri c_{ij}, of C. Furthermore, each an be represented as a linear function of particular input *a priori* and *conditional* probabilities. Thus, if we want all higher order *joint* probabilities to have consistent bounds, we have to find such values for certain of the input probabilities, which are solutions to a system of simultaneous inequalities - each inequality representing the consistency condition for the bounds of each particular higher order *joint* probability.

IV Reasoning with consistent probabilities

The possibility to obtain consistent *bounds* for an arbitrary set of higher order *joint* probabilities helps in determining the probability of *conjunctions* of any number of E. 's; the probability of *disjunctions* of any number of E. 's, and also higher order *conditional* probabilities For IF - THEN rules which premises and/or conclusions are *conjunctions* of an arbitrary number of E_i's.

However, when determining the probability of an arbitrary compound logical proposition or a *conditional* probability consisting of such propositions, an arbitrary set of higher order *joint* probabilities is simply not enough. In this special case one needs higher order *joint* probabilities for the following set of $N = 2^n$ *mutually exclusive events:*

	E_1	E_2	E_3	...	E_{n-2}	E_{n-1}	E_n
P_1	1	1	1	...	1	1	1
P_2	1	1	1	...	1	1	0
P_N	0	0	0	...	0	0	0

Here, 1 expresses that an *event* occurs and 0 that it does not occur. The rows represent a *complete set of mutually exclusive atomic events* where P. is the probability of the i-th of those *events* and $\sum_{i=1:N} P_i = 1$. Now applying the technique proposed in the previous section we can obtain consistent bounds for each *atomic event* P. that confine its actual value. Obtaining consistent probabilities for 2 *complete and mutually exclusive events,* guarantees that we can always find a concrete single value for each of the P. 's, so that they sum up to 1 though, one can as well use the intervals confining it. Then, having assigned such a single value (or an interval) to each *atomic event* we can, as shown in (Konolige 1982), express any logical formula as a disjunction of some subset of (13). Furthermore, the probability of the proposition of interest can be determined by simply summing the probabilities of the corresponding *mutually exclusive atomic events,* members of the disjunction. One advantage that comes from this representation in terms of higher order *joint* probabilities is that loops in the inference net does not matter - the algorithm through which these *joint* probabilities are obtained does not make any use of the concept of directionality.

V Conclusion

The paper presents a new method for computing consistent probabilities for arbitrary logical propositions the main advantages being that: (i) it allows the expert to assess in a consistent way a minimal amount of input-data in terms of *a priori* and *conditional* probabilities; (ii) the amount of computational effort for determining higher order *joint* probabilities is much less when compared to methods based on the minimum-information assumption (Konolige 1982) and (Cheeseman 1983), since in our case only systems of linear inequalities are considered, and (iii) inconsistencies in the intermediate and/or final results, are traced back to the input data and, thus can be resolved by introducing direct changes in some of the input probabilities.

References

1. Kounias, G. (1968), " Bounds for probability of an union with applications." The Annals of Mathematical Statistics 39:6 (1968) 2154-2158.
2. Konolige, K., " An information-theoretic approach to subjective Bayesian inference in rule-based systems." SRI Working Paper, SRI International, Menlo Park,California, March 1982.
3. Cheeseman, P., "A method of computing generalized Bayesian probability values for expert systems." In Proc. IJCAI-83. Karlsruhe, West Germany, August 1983, pp. 198-202.