Probabilistic Knowledge Processing and Remaining Uncertainty

E. Reucher and F. Kulmann
Department of Operations Research
University of Hagen
58084 Hagen, Germany
elmar.reucher@fernuni-hagen.de

Abstract

Information is indispensable in preparing economic decisions purposefully. In this paper knowledge is represented by a probability distribution. Knowledge acquisition is realized by the principle of maximum entropy which guarantees an unsophisticated expansion of knowledge and provides an unbiased basis for decision processes. Under a mathematical point of view entropy measures remaining uncertainty and reduction of entropy is knowledge increase (first order uncertainty). In general, an actual epistemic state might vary due to further information. Thus, it is necessary to have an instrument which reflects also the potential of distribution’s possible variation. Therefore, we present the concept of so called second order uncertainty, which allows not only a ‘look’ on an epistemic state, as the entropy does, but also ‘looks’ into an epistemic state. First the arguments will be complemented by a short example and then they will be illustrated by a real world model in the field of business to business transactions. All calculations will be supported by the expert system SPIRIT.

Introduction

Knowledge processing in a conditional and probabilistic environment under maximum entropy (MaxEnt) and minimum relative entropy (MinREnt), respectively, is a promising approach to support decisions under uncertainty. For axiomatic justifications of these principles the reader is referred to (Csiszár 1975), (Shore and Johnson 1980). The idea is to process incoming facts about the probabilistic conditional structure on all involved variables of a domain in such a way, that not intended dependencies are avoided. The next section gives preliminaries about conditionals and the principle of knowledge processing by entropy. An answer might vary, however, with future knowledge acquisition and hence is subject to second order uncertainty, what Section 'Characterizing a Knowledge Base By Information Measures’ is about. The theoretical concept will be illustrated by an example and a real world application. The paper ends with a conclusion and gives a sketch of further research activities on the field of knowledge processing by entropy.

Preliminaries

Conditionals, Syntax and Semantics

This section gives a theoretical background about conditionals and the principle of knowledge processing under maximum entropy. For a more detailed description the reader is referred to (Roedder 2003), (Roedder and Kern-Isberner 2003). Let \( V = \{ V_1,...,V_N \} \) be a finite set of finite valued variables with respective values \( v_j \) of \( V_j \) and \( v=\{v_1,...,v_N \} \) a configuration which contains exactly one state from each variable of \( V \). Propositions of the form \( V_j = v_j \) are called literals which can be (true) or \( \text{f(alse)} \) for an object or an individuum. From such literals, propositions are formed by the junctors \( \wedge \) (and), \( \vee \) (or), \( \neg \) (not) and by respective parentheses; they are denoted by upper case letters \( A, B, C,... \). Such propositions build the propositional language \( \mathcal{L} \). With the binary conditional operator \( \mid \) formulas of \( B\mid A \) are called conditionals or rules; they build the language \( \mathcal{L}\mid \mathcal{L} \). More on that the reader might find in (Calabrese 1991), (Roedder and Kern-Isberner 2003). By assignment of probabilities \( x \) to a conditional \( B\mid A \), \( x \) expresses the degree to which \( B\mid A [x] \) is true. Such probabilities \( x \) can be given by an expert or can be the result of a dataminings process. For more details c.f (Reucher 2002), (Kern-Isberner and Reidmacher, 1996).

Example:

Given the propositions (to be) \( \text{Student} \) and \( \text{Young} \). The conditional \( \text{Student} \mid \text{Young} [.3] \) means that 30 % of young (people in a fictive population) are students.

As to the semantics, a model is a probability distribution \( P \) on \( \mathcal{L}\mid \mathcal{L} \) that represents the knowledge on a domain, given a set of rules \( \mathcal{R} \).

Knowledge Processing by Entropy

Knowledge Acquisition

Knowledge processing by entropy is a well-known and an axiomatically verified principle (Csiszár 1975), (Shore and Johnson 1980). Given a set of rules \( \mathcal{R} = \{ B_i\mid A_i, [x_i], i = 1, \ldots , I \} \), the knowledge acquisition process is realised by solving the nonlinear optimisation problem

\[
P^* = \arg \min R(Q, P^0), \quad s.t. \ Q \models \mathcal{R}, \tag{1}
\]

Copyright © 2007, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.
where

\[ R(Q, P^0) = \sum_v Q(v) \cdot ld \frac{Q(v)}{P^0(v)} \]

In conformance to former publications on the field of entropy based knowledge acquisition, the notation \( R(Q, P^0) \) is used instead of the in the field of AI well known Kullback–Leibler (KL) divergence KL(\( Q || P^0 \)). \( ld \) denotes the logarithm with bases two, so the quantity of \( R \) has the dimension bit. Here the arg-function determines \( P^* \) as the special probability distribution among all \( Q \) which minimizes \( R(Q, P^0) \) satisfying the linear equations \( R \). \( P^* \) is considered the epistemic state from which all valid conditionals can be evaluated. \( P^0 \) denotes the uniform distribution and is the solution of (1) for \( R = \{ \} \).

A set of conditionals \( R = \{ B_i \mid A_i \mid x_i \}, i = 1, \ldots, I \) represents a convex polyhedron (Reucher 2002). \( R \) is the relative entropy of \( Q \) with respect to \( P^0 \). (1) is equivalent to maximizing the entropy (Roeder 2003):

\[ P^* = \text{arg max } H(Q), \quad \text{s.t. } Q \models R, \quad (2) \]

where

\[ H(Q) = -\sum_v Q(v) \cdot ld Q(v) \]

It is well known, that \( H \) measures the average of uncertainty of any \( v \) being true.

### Query and Response

A focus is a temporary condition, which can be expressed by a rich set of probabilistic rules \( \mathcal{E} = \{ F_j \mid E_j \mid y_j \}, j = 1, \ldots, J \}. \) Such rules provide temporary assumptions about the domain. The adoption of \( P^* \) to the focus \( \mathcal{E} \) yields solving equation

\[ P^{**}(D) = z \quad (4) \]

is the response to this question. Note, that the basic rules \( R \) are adapted to the temporary situation to \( \mathcal{E} \), so that \( R|\mathcal{E} = \{ B_i \mid A_i \mid \tilde{x}_i \}, i = 1, \ldots, I \} \) with \( \tilde{x}_i \neq x_i \) for some \( i \), in general. This will now be illustrated by the well known example Lea Sombé (Sombé 1990).

### Lea Sombé

The knowledge domain consists of three boolean variables Young, Student, Parent and the three-valued nominal variable Marital status (single, married, shared household). The facts about a fictitious population are

\( R = \{ \text{Young \mid Student}[.30], \text{Young \mid Marital}=s[.70], \text{Young \mid Marital}=h[.80], \text{Marital}=s\mid (\text{Student} \land \text{Parent})[.1], \text{Marital}=s\mid \text{Young}[.80] \} \).

Suppose, we are interested in answering the question ‘What’s the probability of someone being Young?’, given the rules \( R \). With \( D = \text{Young} \), the answer is \( P^*(D) = .59 \), where \( P^* \) is the solution of (2) and \( R \) the set of rules, given above. Let us assume a special scenario ‘all young people are single’. \( \mathcal{E} = \{ \text{Marital} = s \mid \text{Young} = 1 \} \), for instance. Then \( P^{**}(D) = .74 \) and the probabilities of some rules in \( R \) are unmodified but differs for some as follows:

\[ R|\mathcal{E} = \{ \text{Young} \mid \text{Student}[.93], \text{Young} \mid \text{Marital}=s[.1], \text{Marital}=s\mid (\text{Student} \land \text{Parent})[.09] \} \]

### Characterizing a Knowledge Base by Information Measures

#### First Order Uncertainty

Given a set of rules \( R \) and \( P^* \) as the solution of (1), (2) respectively. Then the difference \( H(P^0) - H(P^*) \) measures the reduction of the average uncertainty in \( P^* \) from \( P^0 \) by learning \( R \). This difference of entropies is equivalent to knowledge increase which is measurable precisely in [bit]. Refer to Lea Sombé we get \( H(P^0) = 4.58 \) bit and \( H(P^*) = 3.53 \) bit, so that learning the conditionals \( R \) causes knowledge increase of 1.05 bit. It shows, that the reduction of uncertainty (= knowledge increase) is measurable in [bit], precisely.

The entropy is a global measure, because it characterizes a distribution by its inherent remaining average uncertainty. But this characterization may not be sufficient such as in solving decision problems under incomplete information. In particular, a decision maker is more interested in the confidence of answers influencing his decision. In the following section we present a suitable instrument for this.

#### Second Order Uncertainty

Suppose \( D \) be a literal. The answer \( P^*(D) \) might vary, however, with future knowledge acquisition and hence is subject to second order uncertainty. This form of uncertainty allows a further characterization of a knowledge base (Reucher 2006). To calculate the second order uncertainty for an answer \( D \), the following problems have to be solved.

\[ \overline{m} = \min Q(D) \quad \text{s.t. } Q \models R \quad (5) \]

\[ \underline{m} = \max Q(D) \quad \text{s.t. } Q \models R. \quad (6) \]

\( \overline{m}, \underline{m} \) is the interval in which \( Q(D) \) might vary, due to any future knowledge increase \( R := R \cup R' \). Second order uncertainty can be measured by

\[ m = -ld(\overline{m}) - (-ld(\underline{m})). \quad (7) \]

In the light of information theory, \( m \) measures the variability of the estimator \( P^*(D) \) in [bit]. The range of values for \( m \) is \([0; \infty) \). \( m = 0 \) holds iff \( \overline{m} = \underline{m} \) and \( m = \infty \) iff \( \overline{m} = 0 \). For a detailed discussion of this subject see (Roeder 2003), and (Roeder, Reucher, Kulmann, 2006). The equations (5) and (6) can be solved by an algorithm which is implemented in the expert system shell SPIRIT (Spirit 2006).
This theoretical concept is illustrated in continuation of the example Lea Sombé.

**Continuation of Lea Sombé**

Given \( D \equiv \text{YOUNG} \) before, the answer was \( P^*(D) = .59 \). But how precise is this information? Solving (5) and (6) yields [0 : .75]. So, the answer that someone in the population is young may vary between 0% and 75% due to further information given \( R \). Thus, the second order uncertainty \( m \) subject to (7) amounts to \(-l_d(0.) - (-l_d(.75)) = \infty \) bit.

**Expansion of Knowledge**

Now we are interested in expanding a given knowledge base represented by a set of rules \( R \) to \( R \cup R' \). As mentioned in the last chapter, \( R \cup R' \) represents a convex polyhedron, too. If we are considering literal \( D \) given before, it is possible to calculate the interval \([\vec{m} : \vec{m}']\) by solving (5) and (6), respectively. Obviously \([\vec{m} : \vec{m}'] \subset [\vec{m} : \vec{m}']\), because the convex polyhedron spanned by \( R \) can never exceed by the linear restrictions formulated in \( R' \) and will be shortened in general. We now modify the knowledge base 'Lea Sombé' by expanding the set of rules \( R \) to \( R \cup R' \) whereby \( R' = \{ \text{YOUNG} | (\text{STUDENT} \lor \text{Marital} = m) \} \).

Solving the nonlinear optimisation problem (1), we get the solution \( P^* \) and \( H(P^*) = 2.14 \) bit. The difference \( H(P^*) - H(P'^*) \) is the knowledge increase, it amounts to 1.39 bit. Furthermore we are able to analyse second uncertainty for any literal \( D \) again. For \( D \equiv \text{YOUNG} \) the results of (5) and (6) yields \([\vec{m} : \vec{m}'] = [.11 ; .14]\). The second order uncertainty amounts to \( m' = -l_d(.11) - (-l_d(.14)) \) = .35 bit. A comparison between this value \( m' \) and the previous calculated \( m \) shows a significant reduction of second order uncertainty, namely from \( \infty \) to .35 bit. Hence, the answer of \( D \) given \( R \cup R' \) is much more trustworthy than before.

The meaning of first and second order uncertainty measures becomes clear by solving decision problems under incomplete information in real situations. That is what the next section is about.

**An Application with Economic Reference**

**Knowledge Base 'Subcontractor'**

Based on a scoring model, a producer (decision maker) has to decide whether to place an order to a subcontractor (supplier) or not. More precisely, his decision depends on the value of supplier’s profile, which is determined by the following 20 variables:

- A super-criterion denoted by (the supplier’s) Assessment (good, bad). This is the decision relevant variable, which is influenced by
- Four sub-criteria
  - Quality (good, bad),
  - Finance (good, bad),
  - Delivery (good, bad) and
  - Others (good, bad). Each of this criteria itself is specified by so called profilevariables, listed in Table 1.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>TQ, SE, IC, SA, SF, RE</td>
<td>(good, bad)</td>
</tr>
</tbody>
</table>

All profile-variables are also binary with values TQ, SE, IC, SA, SF, RE (good, bad).

**Table 1: Variables of the domain.**

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Finance</th>
</tr>
</thead>
<tbody>
<tr>
<td>QW, PR, SS, CA (high, low),</td>
<td>-Price (PR),</td>
</tr>
<tr>
<td>SC, DC, AC, BBT (yes, no) and</td>
<td>-Discount (DC),</td>
</tr>
<tr>
<td>PR, AR (strict, fair).</td>
<td>-Supplier’s Credit (SC).</td>
</tr>
</tbody>
</table>

The evaluation of a supplier results from scores, given as follows.
1. The score to each of the four sub-criteria is the aggregation of scored corresponding profilevariables.
2. The score of the super-criterion results from aggregating of the scored sub-criteria.

This real scoring model was transformed into a probabilistic equivalent in the form of syntax and semantics presented in the preliminaries of this paper. For a detailed discussion the reader is refered to (Kulmann 2002).

**The Knowledge Bases**

For an illustration of the theoretical aspects, described in the last section, the knowledge base is built up in two steps. In each one, the meaning of first and second order uncertainty’s values is exemplified.

**(1st step)**

Let us assume, that no information about the conditional dependencies between the sub-criteria Quality and Delivery and their corresponding profilevariables is known, so that the knowledge base \( R \) consists of 56 rules. A part of \( R \) is visualised in Figure 1. The rules no. 0 to no. 15 express the relations between the super-criterion Assessment and the four sub-criteria. Read the first rule, for instance: If a supplier’s profile is good for each sub-criteria Quality, Finance, Delivery and Others then SPIRIT answers Assessment = good is true with probability = 1. The interpretation of the other rules is similar. The rules no. 16 to no.
56 describe the probabilistic dependencies between the sub-criteria Finance and Other to their corresponding profile variables.

The model’s hierarchy shows Figure 2 as a dependency graph, which was built in SPIRIT (Spirit 2006).

Each variable is represented by a node. The variable Assessment is displayed at the top, below the four sub-criteria Quality, Finance, Delivery and Others as well as the 15 profile variables. Two nodes $V_i$ and $V_m$ are linked only if:

- in some conditional $B_i|A_i$, a value $v_i$ is involved in $A_i$ and $v_m$ in $B_i$ or
- the values $v_i$ and $v_m$ appear in the conclusion $B_i$ of the same conditional, some $i$.

Figure 1: Excerpt of the knowledge base $\mathcal{R}$.

Figure 2: Model’s hierarchy.

But how trustworthy is this answer? Solving (5) and (6) with $D \equiv \text{Assessment} = \text{good}$ yields the interval of second order uncertainty $\left[\frac{\ln 0.0}{\ln 1.0}\right] = 0.0 \ldots 1.0$. For this, the answer $P^*(\text{Assessment} = \text{good})$ still might vary due to lack of information between 0.0 and 1.0 beyond the basic rules $\mathcal{R}$. Second order uncertainty calculates to $m = -\ln 0.0 - (-\ln 1.0) = \infty$ bit.

Suppose a special supplier with the following profile $E$:

- Quality-profile: $\text{TQ}=\text{good} \land \text{SE}=\text{bad} \land \text{IC}=\text{good} \land \text{QW}=\text{high}$ [1].
- Finance-profile: $\text{PR}=\text{low} \land \text{SC}=\text{yes} \land \text{DC}=\text{no}$ [1].
- Delivery-profile: $\text{SA}=\text{bad} \land \text{AR}=\text{strict} \land \text{SS}=\text{high}$ [1].
- Others-profile: $\text{CA}=\text{high} \land \text{tt}=\text{bad} \land \text{AC}=\text{yes} \land \text{RE}=\text{good} \land \text{BBT}=\text{no}$ [1].

For $D \equiv \text{Assessment} = \text{good}$ solving (3) the system’s answer is $P^{**}(D) = P^*(D|\mathcal{E}) \approx 0.53$. To take the value of second order uncertainty into consideration $[-0.0; 1.0]$, this answer can only be a rough estimation for supplier’s assessment. For this, it is necessary to look for further information about the knowledge domain. This is what step 2 is about.

(2nd step)

Suppose, the decision maker gets more information about the conditional dependencies between the variables of the domain in form of a further set $\mathcal{R}'$, that consists on 44 rules. An excerpt gives Figure 3.
With the additional rules in Figure 3, the conditional structure of the model becomes much more complex. That is visualised by further links between the corresponding variables in Figure 4.

Solving (1) with the extended set of rules \( \mathcal{R} \cup \mathcal{R}’ \) we get \( P^{*’} \) that contains a remaining uncertainty of \( H(P^{*’}) = 18.97 \) bit and the value of knowledge increase amounts to further \( H(P^{*}) - H(P^{*’}) = 19.90 - 18.97 = .33 \) bit. Moreover, the value of second order uncertainty for the answer \( P^{*’}(\text{Assessment}=\text{good}) \) reduces to \( m = -\text{ld}(0.108) - (\text{ld}(0.969)) = 3.155 \) bit significantly in comparison to its value of \( \infty \) bit given the basic rules \( \mathcal{R} \) before. For the supplier with the profile \( \mathcal{E} \) the systems’ answer is \( P^{*’}(D) = .56 \). This answer does not differ very much from the answer before \( (P^{*}(D) = .53) \), but it is much more precised due to two effects:

1. First order uncertainty (= entropy) decreases of .33 bit.
2. Answer’s second order uncertainty decreases from \( \infty \) bit to 3.155 bit.

Table 2 repeats the main results compactly. The second column reflects the result for \( \mathcal{R} = \{ \} \). That means, no information about probabilistic dependencies of the domain’s variables is known. In this case, the uniform distribution \( P^0 \) represents the knowledge base. Because of the domain consists on 20 binary variables, the first order uncertainty (= entropy) of the epistemic state \( P^0 \) is maximum with \( H(P^0) = \text{ld}(\Omega) = 20 \) bit. The second order uncertainty of \( \text{Assessment} = \text{good} \) is also maximum \( m = \infty \) bit. After learning the rules \( \mathcal{R} \) (Figure 1), the uncertainty in the actual epistemic state \( P^* \) decreases to \( H(P^0) - H(P^*) = .7 \) bit, but \( m \) the answers’ uncertainty is still \( \infty \) bit. This shows, that the difference of entropy is a global dimension, whereby the the second order uncertainty is a local dimension.

In a further step, after learning the enriched set of rules \( \mathcal{R} \cup \mathcal{R}’ \) (Figure 3) the additional knowledge increase measures \( H(P^*) - H(P^{*’}) = .33 \) bit. Moreover, the answer’s uncertainty \( \text{Assessment} = \text{good} \) reduces significantly from \( \infty \) bit to 3.155 bit. With these results the decision maker gets sufficient information about the decision situation under incomplete information. Even though the prediction of a supplier with profile \( \mathcal{E} \) is approximately .5 given each of the knowledge bases \( \{ \} \), \( \mathcal{R} \), \( \mathcal{R} \cup \mathcal{R}’ \), but the decision situation is quite different. In particular, under \( \mathcal{R} \cup \mathcal{R}’ \) the answer is most trustworthy!

**Conclusion**

Dealing with uncertainty in the environment of a probabilistic knowledge acquisition process makes it necessary to develop measures characterizing the actual knowledge state. In this paper, the authors presented the measurements first and second order uncertainty which are suitable for characterizing a decision problem under incomplete information. Both concepts are theoretically well-founded and they are implemented in the expert system shell SPIRIT. It allows a computer based calculation of each measure in milliseconds also for more complex models as presented in the last section. Therefore, SPIRIT is a suitable tool for analyzing decision problems under incomplete information.

The examples in this paper contained only consistent knowledge bases, so that (1) always converges to a solution \( P^* \). But SPIRIT is also able to handle inconsistent rules. In this case a so called "inconsistent algorithm" dissolves the rules’ contradiction. Given a set of inconsistent rules \( \mathcal{R} = \{ B_i | A_i[x_i] \} \), the solution of the revision process is a set of consistent rules \( \mathcal{R} = \{ B_i | A_i[x_i] \} \), whereby at least one of \( x_i \) differs from its corresponding \( x_i \). The revision process to \( \pi_i \) bases also on an entropy based algorithm. For more details the reader is refered to (Roedder and Xu 1999).

**References**


