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 (Cszisàr 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 a 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 $\mathbf{v} = \{v_1, ..., v_N\}$ a configuration which contains exactly one state from each variable of V. Prospositions of the form $V_j = v_j$ are called literals which can be t(rue) or f(alse) for an object or an individuum. From such literals, propositions are formed by the junctors $\wedge$ (and), $\vee$ (or), $^-$ (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 $|$ formulas of $B|A$ are called conditionals or rules; they build the language $\mathcal{L}|\mathcal{L}$. More on that the reader might find in (Calabrese 1991), (Roedder and Kern-Isberner 2003). By assignment of probabilities x to a conditional $B|A$, x expresses the degree to which $B|A[x]$ 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) `Student` and `Young`.
The conditional `Student|Young` [.3] means that 30 % of young (people in a ficitve population) are students.

As to the semantics, a model is a probability distribution P on $\mathcal{L}|\mathcal{L}$ 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 (Csiszàr 1975), (Shore and Johnson 1980). Given a set of rules $\mathcal{R} = \{B_i|A_i[x_i], i = 1,\ldots,I\}$, the knowledge acquisition process is realised by solving the nonlinear optimisation problem

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

where

$$R(Q, P^0) = \sum_{\mathbf{v}} Q(\mathbf{v}) \cdot ld \frac{Q(\mathbf{v})}{P^0(\mathbf{v})}$$

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 $\mathcal{R}$. 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 $\mathcal{R} = \{\}$.

A set of conditionals $\mathcal{R} = \{B_i | A_i [x_i], i = 1, \ldots, I\}$ 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 (Roedder 2003):

$$P^* = \arg\max H(Q), \quad s.t.\ Q \models \mathcal{R}, \tag{2}$$

where

$$H(Q) = -\sum_{\mathbf{v}} Q(\mathbf{v}) \cdot ld\, Q(\mathbf{v})$$

It ist well known, that H measures the average of uncertainty of any $\mathbf{v}$ being true.

Query and Response

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

$$P^{**} = \arg\min R(Q, P^*) \quad s.t.\ Q \models \mathcal{E}. \tag{3}$$

A question might be any literal D, whose probability should be evaluated under basic rules $\mathcal{R}$ and the focus $\mathcal{E}$. A focus $\mathcal{E}$ describes a special scenario.

$$P^{**}(D) = z \tag{4}$$

is the response to this question. Note, that the basic rules $\mathcal{R}$ are adapted to the temporary situation to $\mathcal{E}$, so that $\mathcal{R}|\mathcal{E} = \{B_i | A_i [\tilde{x}_i], i = 1, \ldots, I\}$ 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 (<u>s</u>ingle, <u>m</u>arried, shared <u>h</u>ousehold). The facts about a fictitious population are

$\mathcal{R} = \{$ `Student|Young [.30],`
    `Young|Student [.90],`
    `Young|Marital=s [.70],`
    `Young|Marital=h [.80],`
    `Marital=s|(Student∧ Parent) [.1],`
    `Marital=s | Young [.80]}.`

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

$\mathcal{R}|\mathcal{E} = \{$ `Young|Student [.93],`
    `Young|Marital=s [1.],`
    `Marital=s|(Student∧ Parent) [.09]}.`

Characterizing a Knowledge Base by Information Measures

First Order Uncertainty

Given a set of rules $\mathcal{R}$ 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 $\mathcal{R}$. 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 $\mathcal{R}$ 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 decison 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.

$$\overline{u} = \min Q(D) \quad s.t.\ Q \models \mathcal{R} \tag{5}$$

$$\overline{\overline{u}} = \max Q(D) \quad s.t.\ Q \models \mathcal{R}. \tag{6}$$

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

$$m = -ld(\overline{u}) - (-ld(\overline{\overline{u}})). \tag{7}$$

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 $\overline{u} = \overline{\overline{u}}$ and $m = \infty$ iff $\overline{u} = 0$. For a detailed discussion of this subject see (Roedder 2003), and (Roedder, 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 $D \equiv$ 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 $\mathcal{R}$. Thus, the second order uncertainty m subject to (7) amounts to $-ld(0.) - (-ld(.75)) = \infty$ bit.

Expansion of Knowledge
Now we are interested in expanding a given knowledge base represented by a set of rules $\mathcal{R}$ to $\mathcal{R} \cup \mathcal{R}'$. As mentioned in the last chapter, $\mathcal{R} \cup \mathcal{R}'$ represents a convex polyhedron, too. If we are considering literal D given before, it is possible to calculate the interval $[\overline{u}'; \overline{\overline{u}}']$ by solving (5) and (6), respectively. Obviously $[\overline{u}'; \overline{\overline{u}}'] \subseteq [\overline{u}; \overline{\overline{u}}]$, because the convex polyhedron spanned by $\mathcal{R}$ can never exceed by the linear restrictions formulated in $\mathcal{R}'$ and will be shorten in general. We now modify the knowledge base 'Lea Sombé' by expanding the set of rules $\mathcal{R}$ to $\mathcal{R} \cup \mathcal{R}$' whereby $\mathcal{R}'=\{$Young $| \overline{(\text{Student}} \vee \text{Marital=m})$ $[.1]\}$. 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 $D \equiv$ Young the results of (5) and (6) yields $[\overline{u}'; \overline{\overline{u}}'] = [.11; .14]$. The second order uncertainty amounts to $m' = -ld(.11) - (-ld(.14)) = .35$ bit. A comparison between this value m' and the previous calculated m shows a significant reduction of second order uncertainty, namely from ∞ to $.35$ bit. Hence, the answer of D given $\mathcal{R} \cup \mathcal{R}$' 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 decisison relevant varaible, 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.

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

Assessment	
Quality	Finance
-Technical Quality (TQ) -Quote of Waste (QW) -Image of Client (IC) -Service (SE)	-Price (PR) -Discount (DC) -Supplier's Credit (SC)

Delivery	Others
-Supplier's Accuracy (SA) -Supplier's Security (SS) -Attitude by Reclamation (AR)	-Attendance to -Cooperation (AC) -B2B Transactions (BBT) -Capacity (CA) -Reputation (RE) -Satisfaction of Client (SF)

Table 1: Variables of the domain.

QW, PR, SS, CA (high, low),
SC, DC, AC, BBT (yes, no) and
PR, AR (strict, fair).

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 $\mathcal{R}$ consists of 56 rules. A part of $\mathcal{R}$ 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.

Index	P act	Rule text
0	1,00000	Assessment=good | Quality=good ∧ Finance=good ∧ Delivery=good ∧ Others=good
1	0,75000	Assessment=good | Quality=good ∧ Finance=good ∧ Delivery=good ∧ Others=bad
2	0,00000	Assessment=good | Quality=bad ∧ Finance=bad ∧ Delivery=bad ∧ Others=bad
3	0,75000	Assessment=good | Quality=good ∧ Finance=good ∧ Delivery=bad ∧ Others=good
4	0,50000	Assessment=good | Quality=good ∧ Finance=good ∧ Delivery=bad ∧ Others=bad
5	0,75000	Assessment=good | Quality=good ∧ Finance=bad ∧ Delivery=good ∧ Others=good
6	0,50000	Assessment=good | Quality=good ∧ Finance=bad ∧ Delivery=good ∧ Others=bad
7	0,50000	Assessment=good | Quality=good ∧ Finance=bad ∧ Delivery=bad ∧ Others=good
8	0,25000	Assessment=good | Quality=good ∧ Finance=bad ∧ Delivery=bad ∧ Others=bad
9	0,75000	Assessment=good | Quality=bad ∧ Finance=good ∧ Delivery=good ∧ Others=good
10	0,50000	Assessment=good | Quality=bad ∧ Finance=good ∧ Delivery=good ∧ Others=bad
11	0,50000	Assessment=good | Quality=bad ∧ Finance=good ∧ Delivery=bad ∧ Others=good
12	0,25000	Assessment=good | Quality=bad ∧ Finance=good ∧ Delivery=bad ∧ Others=bad
13	0,50000	Assessment=good | Quality=bad ∧ Finance=bad ∧ Delivery=good ∧ Others=good
14	0,25000	Assessment=good | Quality=bad ∧ Finance=bad ∧ Delivery=good ∧ Others=bad
15	0,25000	Assessment=good | Quality=bad ∧ Finance=bad ∧ Delivery=bad ∧ Others=good
16	0,00000	Finance=good | PR=high ∧ SC=no ∧ DC=no
17	0,17000	Finance=good | PR=high ∧ SC=no ∧ DC=yes
18	0,33000	Finance=good | PR=high ∧ SC=yes ∧ DC=no
19	0,50000	Finance=good | PR=high ∧ SC=yes ∧ DC=yes
20	0,50000	Finance=good | PR=low ∧ SC=no ∧ DC=no
21	0,67000	Finance=good | PR=low ∧ SC=no ∧ DC=yes
22	0,83000	Finance=good | PR=low ∧ SC=yes ∧ DC=no
23	1,00000	Finance=good | PR=low ∧ SC=yes ∧ DC=yes
24	0,00000	Others=good | SF=bad ∧ RE=bad ∧ CA=low ∧ BBT=no ∧ AC=no
25	0,11000	Others=good | SF=bad ∧ RE=bad ∧ CA=low ∧ BBT=yes ∧ AC=no
⋮		
53	0,89000	Others=good | SF=good ∧ RE=good ∧ CA=high ∧ BBT=no ∧ AC=yes
54	0,66500	Others=good | SF=good ∧ RE=good ∧ CA=high ∧ BBT=yes ∧ AC=no
55	1,00000	Others=good | SF=good ∧ RE=good ∧ CA=high ∧ BBT=yes ∧ AC=yes

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

56 describe the probabilistic dependencies between the sub-criteria `Finance` and `Other` to their corresponding profilevariables.

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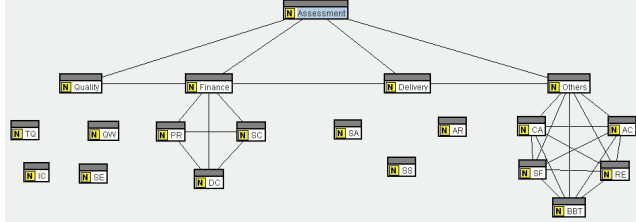
Figure 2: Model's hierarchy.

Each variable is represented by a node. The variable `Assessment` is displayed at he top, below the four subcriteria `Quality`, `Finance`, `Delivery` and `Others` as well as the 15 profilevariables. Two nodes V_l and V_m are linked only if

- in some conditional $B_i|A_i$ a value v_l is involved in A_i and v_m in B_i or

- the values v_l and v_m appear in the conclusion B_i of the same conditional, some i.

Suppose P^* is the solution of (1) and $\mathcal{R}$ denotes the set of 56 rules described above. The knowledge base remains $H(P^*) = 19.3$ bit (first order) uncertainty and the knowledge increase amounts to $H(P^0) - H(P^*) = 20.0 - 19.3 = .7$ bit. SPIRIT gives the answer for any producer of 'being good' with an estimation of $P^*(\texttt{Assessment = good}) = .39$.

But how trustworthy is this answer? Solving (5) and (6) with $D \equiv \texttt{Assessment = good}$ yields the interval of second order uncertainty $[\overline{u}; \overline{\overline{u}}] = [0.0; 1.0]$. For this, the answer $P^*(\texttt{Assessment = good})$ still might vary due to lack of information betweeen 0.0 and 1.0 beyond the basic rules $\mathcal{R}$. Second order uncertainty calculates to $m = -ld0.0 - (-ld1.0) = \infty$ bit.

Suppose a special supplier with the following profile $\mathcal{E}$:

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

For $D \equiv \texttt{Assessment = good}$ solving (3) the system's answer is $P^{**}(D) = P^*(D|\mathcal{E}) \approx .53$. 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.

Index	P act	Rule text
56	0,12500	Quality=good | TQ=bad ∧ QW=high ∧ IC=good ∧ SE=bad
57	0,25000	Quality=good | TQ=bad ∧ QW=high ∧ IC=bad ∧ SE=good
58	0,37500	Quality=good | TQ=bad ∧ QW=high ∧ IC=good ∧ SE=good
59	0,00000	Quality=good | TQ=bad ∧ QW=high ∧ IC=bad ∧ SE=bad
60	0,25000	Quality=good | TQ=bad ∧ QW=low ∧ IC=bad ∧ SE=bad
61	0,50000	Quality=good | TQ=bad ∧ QW=low ∧ IC=bad ∧ SE=good
62	0,37500	Quality=good | TQ=bad ∧ QW=low ∧ IC=good ∧ SE=bad
63	0,62500	Quality=good | TQ=bad ∧ QW=low ∧ IC=good ∧ SE=good
64	0,37500	Quality=good | TQ=good ∧ QW=high ∧ IC=bad ∧ SE=bad
65	0,62500	Quality=good | TQ=good ∧ QW=high ∧ IC=bad ∧ SE=good
66	0,50000	Quality=good | TQ=good ∧ QW=high ∧ IC=good ∧ SE=bad
67	0,75000	Quality=good | TQ=good ∧ QW=high ∧ IC=good ∧ SE=good
68	0,62500	Quality=good | TQ=good ∧ QW=low ∧ IC=bad ∧ SE=bad
69	0,87500	Quality=good | TQ=good ∧ QW=low ∧ IC=bad ∧ SE=good
70	0,75000	Quality=good | TQ=good ∧ QW=low ∧ IC=good ∧ SE=bad
71	1,00000	Quality=good | TQ=good ∧ QW=low ∧ IC=good ∧ SE=good
72	0,50000	QW=low | TQ=good
73	0,50000	QW=low | TQ=bad
74	0,50000	IC=good | QW=low
75	0,50000	IC=good | QW=high
76	0,50000	SE=good | IC=good
77	0,50000	SE=good | IC=bad
78	0,50000	TQ=good | SE=good
79	0,50000	TQ=good | SE=bad
80	0,00000	Delivery=good | SA=bad ∧ SS=low ∧ AR=strict
⋮		
96	0,50000	DC=yes | SC=yes
97	0,50000	DC=yes | SC=no
98	0,50000	PR=low | DC=yes
99	0,50000	PR=low | DC=no

Figure 3: Exceprt of the knowledge base $\mathcal{R}'$.

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.
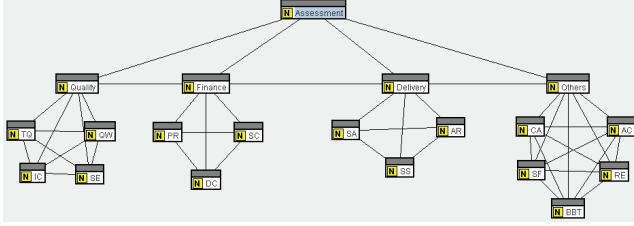

Figure 4: Extended Model.

Solving (1) with the extended set of rules $\mathcal{R} \cup \mathcal{R}'$ we get $P^{*\,\prime}$ that contains a remaining uncertainty of $H(P^{*\,\prime})$=18.97 bit and the value of knowledge increase amounts to further $H(P^*) - H(P^*)' = 19.30 - 18.97 = .33$ bit. Moreover, the value of second order uncertainty for the answer $P^{*\,\prime}$(Assessment=good) reduces to $m = -ld(0.108) - (-ld(.969)) = 3.155$ bit significantly in comparison to its value of ∞ bit given the basic rules $\mathcal{R}$ before. For the supplier with the profile $\mathcal{E}$ the systems' answer is $P^{*\,\prime}(D) = .56$. 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 ∞ bit to 3.155 bit.

Table 2 repeats the main results compactly. The second

	$\{\}$ P^0	$\mathcal{R}$ P^*	$\mathcal{R} \cup \mathcal{R}'$ $P^{*\,\prime}$
No. of rules	0	56	100
H	20.00	19.30	18.97
m	∞	∞	3.155
$P(D\|\mathcal{E})$	.50	.53	.56

Table 2: Characterization of knowledge bases

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 $H(P^0) = ld|\Omega| = 20$ bit. The second order uncertainty of Assessment=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 $H(P^0) - H(P^*) = 0.7$ bit, but m the answers' uncertainty is still ∞ 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 $H(P^*) - H(P^{*\,\prime}) = .33$ bit. Moreover, the answer's uncertainty Assessment = good reduces significantly from ∞ bit to 3.155 bit.

With these results the decision maker gets sufficient information about the decision situation under incomplete information. Even though the predicition 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 $\mathcal{R} \cup \overline{\mathcal{R}}$ the answer is most trustworthy!

Conclusion

Dealing with uncertainty in the environment of a probablistic 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 $\overline{\mathcal{R}} = \{B_i|A_i\,[\overline{x_i}]\}$, whereby at least one of $\overline{x_i}$ differs from its corresponding x_i. The revision process to $\overline{x_i}$ bases also on an entropy based algorithm. For more details the reader is refered to (Roedder and Xu 1999).

References

Calabrese, P.G. 1991. Deduction and Inference Using Conditional Logic and Probability, Conditional Logic in Expert Systems. *Elsevier Science Publishers B. V*, 71-100.

Csiszàr, I. 1975. I-Divergence Geometry of Probability Distributions and Minimization Problems.*The Annals of Probability, Vol. 3, No.1*, 148-158.

Kern-Isberner, G.; Reidmacher, H.-P. 1996. Interpreting a Contingency Table by Rules, *International Journal of Intelligent Systems, 11(6)*, 327-346.

Kulmann, F. 2002. Wissen und Information in konditionalen Modellen. *Deutscher Universitaetsverlag*, Wiesbaden.

Reucher, E. 2002. Modellbildung bei Unsicherheit und Ungewißheit in konditionalen Strukturen. *Logos*, Berlin.

Reucher, E. 2006. Discovery of Answers' Uncertainty in Conditional Structures. In *Proceedings of the Information Processing and Management of Uncertainty in Knowledge-Based Systems*, IPMU, 9-13, Paris.

Reucher, E.; Rödder, W. 2001. Relevanz von Information in konditionalen Entscheidungsmodellen. In *Proceedings of Operations Research*, 379-386. Springer, Berlin.

Rödder, W. 2003. On the Measurability of Knowledge Acquisition and Query Processing. In *International Journal of Approximate Reasoning*, 33/2, 203-218.

Rödder, W.; Kern-Isberner, G. 2003. From Information to Probability - An Axiomatic Approach. In *International Journal of Intelligent Systems*, 18/4, 383-403.

Rödder, W.; Reucher, E.; Kulmann, F. 2006. Features of the Expert-Sytem Shell SPIRIT. In *Logical Journal of the IGPL*, Oxford Press, 483-500.

Rödder, W.; Xu, L. 1999. Entropy-driven Inference and Inconsistency, Proc. Artificial Intelligence and Statistics, Fort Lauderdale, 272-279,

Shore, J.E., Johnson, R.W. 1980. Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross Entropy. In *IEEE Trans. Inform. Theory IT*-26, 1, 26-37.

Sombé, L. 1990. Reasoning Under Incomplete Information in Artificial Intelligence: A Comparison of Formalisms Using a Simple Example. *John Wiley & Sons*.

SPIRIT 2006. http://www.xspirit.de (2006-11-01).