ENTROPY-DRIVEN INFERENCE AND INCONSISTENCY

Wilhelm Rödder, Longgui Xu
FB WiWi, LS BWL, insb. OR
FernUniversität Hagen, Germany

Abstract

Probability distributions on a set of discrete variables are a suitable means to represent knowledge about their respective mutual dependencies. When now things become evident such a distribution can be adapted to the new situation and hence submitted to a sound inference process. Knowledge acquisition and inference are here performed in the rich syntax of conditional events. Both, acquisition and inference respect a sophisticated principle, namely that of maximum entropy and of minimum relative entropy. The freedom to formulate and derive knowledge in a language of rich syntax is comfortable but involves the danger of contradictions or inconsistencies. We develop a method how to solve such inconsistencies which go back to the incompatibility of experts' knowledge in their respective branches. The method is applied to the diagnosis in Chinese medicine. All calculations are performed in the Entropy-driven expert system shell SPIRIT.

1. INTRODUCTION

1.1 REFLECTIONS ON ENTROPY AND RELATIVE ENTROPY IN AI

An increasing number of scientists and experts applies probabilistic models in AI. Especially Bayes- or Markov-Nets and their generalisations are widely accepted to build and manipulate distributions which contain knowledge about a certain context. For a deeper analysis of the distributions' properties and their representation in graphical structures the reader is referred to Lauritzen [4] and Whittaker [10], among others. As this knowledge processing requires the estimation of innumerable probabilities, it is somewhat unwieldy in acquisition and furthermore it is poor in response. A new generation of knowledge bases uses entropy as the structuring element instead of DAGs or other graphical devices. Entropy was established by the physicists Carnot and Clausius in thermodynamics. It measures the portion of thermal energy which cannot be transformed back into mechanical energy. The engineer Shannon rediscovered entropy as the average information rate which can be transmitted by an optimal codified alphabet of signals via a channel. Shannon [8] showed this information to be (0 ld 0 = 0):

$$H = -\sum_{v} p(v) \,\mathrm{ld}\, p(v).$$

Here the signals v might vary in a finite alphabet V, $v \in V$. $p(v)$ is the probability of a signal's occurrence and $-\mathrm{ld}\, p(v)$ the received information. ld is the dual logarithm and H measures in [bit]. It is well-known, that H assumes its maximum for equally probable v's and its minimum if for one v^0, $p(v^0)=1$.

Similar to the information theoretical context we now assume $p(v)$ to express the probability that an elementary event v in a population with distribution P becomes true. Then $-\mathrm{ld}\, p(v)$ is again, as reduction of uncertainty, information and H its expected value in P. A high H means high uncertainty and allows of a high average reduction, a low H indicates low uncertainty.

Let us assume now the elementary event to be characterised by a finite number of attributes or values v_l of finite-valued variables V_l:$v = v_1,...,v_n$ such as colour, age, sex, medical symptoms etc. To get a deeper insight in entropy we use the factorisation (0/0 = 0)
$p(v_1,...,v_n) = p(v_1)\; p(v_2|v_1)\dots p(v_n|v_1,...,v_{n-1})$ and after some reordering and summations we receive

$$\begin{aligned} H(P) = &-\sum_{v_1} p(v_1)\,\mathrm{ld}\, p(v_1) - \sum_{v_1} p(v_1) \sum_{v_2} p(v_2|v_1)\,\mathrm{ld}\, p(v_2|v_1) \\ &\dots \\ &- \sum_{v_1,...,v_{n-1}} p(v_1,...,v_{n-1}) \sum_{v_n} p(v_n|v_1,...,v_{n-1})\,\mathrm{ld}\, p(v_n|v_1,...,v_{n-1}). \end{aligned} \tag{1}$$

H measures the mutual indefiniteness of all involved Variables V_l from each other in the distribution P. The less information we get from the conditioning variables about the conditioned attributes the higher is entropy – and this statement is valid for any order of the Variables V_l ! For any permutation of the attributes $v = v_{l_1} \dots v_{l_n}$ the probability $p(v)$ remains unchanged and hence does H. Our explication of entropy is merely intuitive and should be supplemented by the very ambitious axiomatic works [2], [3], [9]. From these contributions, especially [3], the reader will learn also that the relative entropy

$$R(Q,P) = \sum_{v} q(v)\,\mathrm{ld}\, {q(v)}/{p(v)} \tag{2}$$

measures the average change in mutual indefiniteness after the distribution P has been transformed to Q. More on such transformations we shall discuss in section 2.

The importance of entropy we demonstrate in a little example.

Example 1 (Bayes-Nets vs. Entropy)
The only information about the distribution of Balls and Dies with colours white and green in a box is that all Balls are white, not more and not less. There are infinite many distributions reflecting this knowledge, three of which are shown in the following table

Table 1: Distributions for incomplete information

Bw	*Bg*	*Dw*	*Dg*	Entropy [bit]
1/5	0	2/5	2/5	1.522
2/5	0	2/5	1/5	1.522
1/3	0	1/3	1/3	1.585.

The first row results from the fact that the knowledge engineer in the Bayes-Net

Shape $\longrightarrow$ Colour

estimated the probability of an object to be a Ball by 20 % and was indecisive with respect to the colour of Dies. Note that this first assessment results in a high definiteness of colour and of the shape given colour:

$w = 3/5$, $g = 2/5$, $B|w = 1/3$, $D|w = 2/3$, $B|g = 0$, $D|g = 1$.

But the definiteness for white objects is not intended. To see this mind the fact that the statement *Balls are white* is the logical contraposition to *green objects are Dies*. Now in turn the 20 % estimation of objects to be green together with the indecision concerning the shape of white objects yield the conditioned probabilities

$w = 4/5$, $g = 1/5$, $B|w = 1/2$, $D|w = 1/2$, $B|g = 0$, $D|g = 1$

which coincides with the second row in table 1. Note that now in turn the earlier indecision with respect to the colour of Dies vanishes... .The only assessment which avoids such circular confusion, is that of maximal entropy in the third row, cf. the last column.

In the case of incomplete information, the arbitrary assessment of probabilities in Bayes-Nets might cause severe errors, even worse if it serves as a basis for decision making.

1.2 CONDITIONALS AND PROBABILITY

In the remainder of this paper we consider a probability space on the field of all events on the set of finite-valued Variables $\mathbf{V} = \{V_1,...,V_n\}$. Each event can be identified in a natural way with propositional sentences built from literals $V_l = v_l$ and the connectives negation, conjunction and disjunction. The set of all such propositional sentences or events we write as capital letters, indexed if necessary: A, B, C,..., A_i, E_j, G_k. If A is such an event, its probability is $P(A) = \sum_{v \in A} p(v)$, where the summation is taken over all elementary events in A. For two events B and A we call $B|A$ a conditional and for a positive $P(A)$ we define $P(B|A)=P(AB)/P(A)$, i.e. the probability of a conditional is identified with its conditioned probability. A conditional is often called a *rule*, if it is conditioned only by the all-event, we call it a *fact*.
Conditional probabilities are the basis for a sound inference principle, as we develop in the next section. For an extensive justification of conditionals to be the adequate probabilistic pendant to material implication in binary logic we refer the reader to [5].

2. ENTROPY-DRIVEN INFERENCE

Entropy-driven inference consists of three steps: the acquisition of knowledge, the processing of evidence and the evaluation of response. Each of these steps may involve uncertainty.

In the first step the knowledge engineer supplies conditionals $B_i|A_i$ $i = 1,...,I$ for which he is able to estimate their probabilities x_i to be true in the respective population. This knowledge then is implemented in a distribution P^* :

$$P^* = \arg\max H(P) \text{ s.t. } P(B_i|A_i) = x_i \quad i = 1,...,I. \qquad (3)$$

P^* contains only the knowledge supplied by the rules and facts, and what can be deduced from them respecting the principle of maximum entropy. Note that the mutual indefiniteness as in (1), is maximum subject to the desired (conditional) probabilities.

In the second step further hypothetical assumptions about the population give rise to a modification of the probability distribution. If we now assume that this certain or uncertain additional information is evident in the form of further conditionals $F_j|E_j$ and their respective probabilities y_j, then we can calculate

$$P^{**} = \arg\min R(P, P^*) \text{ s.t. } P(F_j|E_j)=y_j \quad j = 1,...,J. \qquad (4)$$

P^{**} preserves the indefiniteness of P^* as far as possible, but adapts it to the evident or hypothetical facts and rules. This inference process respects the principle of minimum relative entropy.

In the third step we evaluate the probabilities of conditional questions which the user may have asked the system

$$P^{**}(H_k|G_k) \qquad k = 1,..., K.$$

Such a $P^{**}(H_k|G_k)$ is the answer about the probability of H_k given G_k derived from the knowledge immanent in P^* *and* from the actual evident situation.

The mathematics of this process we presented two years ago on Uncertainty 96 [6] and do not repeat it now. We only mention that (3) is solved by a generalised form of IPF, applying the linear system (5) to the uniform P_0 .

$$(1-x_i)P(B_i A_i) - x_i P(\overline{B}_i A_i) = 0 \quad i = 1,...,I. \qquad (5)$$

Here barring indicates negation. Note that the equation $(1-x_i)P(B_i A_i) - x_i P(\overline{B}_i A_i) = 0$ for a positive $P(A_i)$ is equivalent to $P(B_i|A_i)=x_i$. (4) is solved similarly.

Csiszár [1] showed in a much more general measure-theoretic approach that this IPF converges to a unique solution P^* if *all* restrictions in (5) are *consistent*. Otherwise we call the rules in (3) *inconsistent*, they are contradictory. We shall treat this case in the next section.
There might be a weaker form of inconsistency in that the equations in (5) are consistent but the unique IPF-solution yields $P^*(A_i) = 0$ for some i. This case we call *conditional inconsistency*. It is weaker because concluding arbitrary propositions from a contradictory premise does not cause any harm, as we know from binary logic.

The following serves as an example for a structure of cyclic rules as well as uncertain evidence.

Example 2 (Cyclic Knowledge Acquisition)
In a population of equally distributed men, Sex = m, and women, SEX = f, we consider smoking behaviour, smoking SM = sm and non-smoking SM = sn. Let the affinity to alcohol be yes, AL = y, and no, AL = n. The cyclic knowledge which an expert might supply, we visualise in the following graph

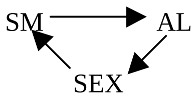

Figure 1: A cyclic graph

The corresponding conditional probabilities are:
(AL = y|SM = sm) = .7, (SEX =m|AL = y) = .7, (SM = sm|SEX = f) = .1, (SM = sm|SEX = m) = .4. Note that this information is, apart from being cyclic, still incomplete.

The Shell SPIRIT [6] iterates this knowledge and is now able to process uncertain evidence. A young man comes in and because of his age we guess with a probability of .95 that he is very likely a smoker, if he drinks alcohol. Does he smoke? The answer is SM = sm with a probability of .64, much more than the .40 among men without the additional uncertain information.

The reader might realise that such cyclic information can cause contradictions. We give an example in the next section and show how to solve them.

3. INCONSISTENCY AND A SOLUTION

3.1 INCONSISTENCY AND A WORLD-VARIABLE

Inconsistency in entropy-driven knowledge bases might occur because of the availability of the language of conditionals in a rich syntax for communication between the user and system. Inconsistency can appear in a weak and in a hard form. In the latter the knowledge provided to the system as conditionals $B_i|A_i \quad i=1,...,I$ and their respective desired probabilities x_i does not allow the construction of a distribution P^*, cf. our argumentation following formula (5).
As a preparation for section 3.2 we now study the modified conditionals $B_i|A_iW$, where W is a binary *world*-variable with values *true* and *false*. Now the modified knowledge acquisition process (3) becomes

$$P^* = \arg\max H(P) \text{ s.t. } P(B_i|A_iW)=x_i \quad i = 1,...,I. \qquad \text{(3W)}$$

We distinguish between P and P because the event field is now enriched by the variable W. Of course, (3W) is solved similar to (3) via generalised IPF now applying to the linear system

$$(1-x_i)P(B_i A_i W) - x_i P(\overline{B}_i A_i W) = 0 \quad i = 1,...,I. \qquad \text{(5W)}$$

Fortunately the relations between (3), (3W) and (5), (5W) can be summarised in the following theorem, the proof of which is straightforward and omitted here.

Theorem (Conditioning World)
i) The problem (3W) is either consistent or conditionally inconsistent, but never inconsistent.
ii) The problem (3) is consistent or conditionally inconsistent, respectively, iff (3W) has the same properties with $P^*(W = \text{t})>0$. In either case the solution P^* of (3) and P^* of (3W) are related by the equation $P^* = P^*(\cdot|W = \text{t})$.
iii) The problem (3) is inconsistent iff (3W) is conditionally inconsistent with $P^*(W = \text{t})=0$.

The theorem shows a strong relationship between (3) and (3W). Roughly speaking, if (3) is inconsistent there is no world W = *true* in which all conditionals hold. Otherwise this world is possible.

3.2 MATCHING INCONSISTENT WORLDS

Section 3.1 served as a first introduction to the concept of knowledge worlds. The present reflections show how to match knowledge of different contexts.

Consider a partition $\mathbf{U} = \{U_1,...,U_K\}$ of $\{1,...,I\}$ and a binary variable W_k for each U_k . Consider furthermore the optimisation problem

$$P^* = \arg\max H(P) \text{ s.t. } P(B_i|A_iW_k) = x_i \quad i \in U_k \text{ all } k \text{ and } P(\wedge W_k | \vee W_k) = \bar{x} \qquad (6)$$

The first restrictions are well known. They guarantee an adequate adaptation of P^* to all established and consistent provinces. The last restriction is parametric in $\bar{x}$. $\bar{x}$ expresses the probability of *all* worlds to be true if only *one* has this property. If $\bar{x}$ is high the worlds W_k condition each other strongly, if it is low they do not. Mind the fact that $\bar{x}$ = 1 shows a perfect mutual dependence of all W_k on each other. All W_k can be substituted by a single *W*, which in turn proves the different knowledge parts to be compatible.

Once we got a good P^* in (6) we can match all partial knowledge as in the following

Definition (Matching of Knowledge)
Let P^* be the solution of (6) and let Q^* be defined as $Q^*=P^*\left(\cdot\,|W_1 = t \wedge ... \wedge W_K = t\right)$. The actualised probabilities $x_i^* = Q^*\left(B_i|A_i\right)$ are called *matching probabilities* and the solution P^* of (3) with x_i^* instead of x_i is a **U-***matching* of (6) at level $\bar{x}$.

From the definition we learn that for a maximum $\bar{x}$, the x_i^* are the "closest" probabilities to x_i. If we look for a situation in which the incompatible information shares would combine best, it is $Q^*=P^*\left(\cdot\,|W_1 = t \wedge ... \wedge W_K = t\right)$, or equivalently

$$Q^* = \arg\min R\left(P, P^*\right) s.t. P\left(W_1 = t \wedge ... \wedge W_K = t\right) = 1. \quad (7)$$

The $x_i^* = Q^*(B_i|A_i)$ then are a compromise between the experts' different meanings. They are the basis for a final knowledge acquisition. P^* from (3) with these conditional probabilities x_i^* obeys the principle of maximum entropy and joins the knowledge of all provinces.

With these considerations we are ready to study a medium-size knowledge base with contradictory rules and facts. In the next section we show an example of diagnosis in Chinese medicine and then apply it to the matching process.

4. DIAGNOSIS IN CHINESE MEDICINE, AN APPLICATION

4.1 THE EIGHT GUIDE PRINCIPLES

Similar as to western medicine, the entire recognition of a disease is also the main challenge to Chinese medicine. A central role plays the syndrome, in Chinese *Zheng*. Syndromes in Chinese medicine involve the origin, place, properties and symptoms of a disease[1]. The basis for a dialectic diagnosis are those syndromes, which Chinese medicine divides in five groups:

- The eight guide principles
- Qi, blood and body liquids
- Internal organs and hallow organs
- The six meridians
- Resistance, Qi, nutrition, blood and the three heaters.

The mutual influence of *all* syndrome-diagnostics is rather complicated. So we decided to analyse one of them and model it in a knowledge base.

Therefore consider the eight guide principles yin/yang, surface/inside, cold/heat, emptiness/abundance. They are dialectic in pairs. In general a cold-syndrome characterises the penetration of a cold-disturbance and a weakness of our organism, whereas a heat-syndrome implies in increasing activities of the metabolism. The connection of such a disturbance to yin/yang can be expressed in two dogmas:

- The abundant yin yields cold; the abundant yang yields heat.
- The empty yang generates cold; the empty yin generates heat.

With emptiness and abundance Chinese doctors characterise the strength or weakness of a patient's resistance. In general an emptiness-syndrome indicates a weak resistance, an abundance-syndrome shows a strong pathogen disturbance.

Surface and inside characterise the place of a disease in human organism. An illness is slight, if it attacked only the surface; it is severe, if it strikes internal and hallow organs.

Yin and yang are the most important of the eight guide principles. All other principles are subordinated to yin and yang. Chinese doctors consider all human psychology and pathology through the dialectic relations between yin and yang, their bipolar and joint influence.

4.2 THE ENTROPY-DRIVEN MODEL AND RESULTS

The aim of this section is to model the knowledge about the eight guide principles by variables and rules and show their relations to the symptom *fever*.

We consider the variables cold/heat: KKH, surface/inside: KOI, emptiness/abundance: KLF, yin/yang: KYY, yin: YIN, yang: YANG, fever: FEVER.

Each of these variables might assume values which we list in the following table

Table 2: Variables and Attributes

KKH	*KOI*	*KLF*	*YIN*	*YANG*
ks	*os*	*ls*	*se*	*se*
hs	*is*	*fa*	*uem*	*uem*
ohuk	*olif*	*gm*	*n*	*n*
okuh	*ofil*	*n*	*vm*	*vm*
ehfk	*okih*	*KKY*	*sm*	*sm*
ekfh	*ohik*	*yis*	*FEVER*	
n	*n*	*yas*	*ef*	*sf*
		n	*mf*	*n*

The meaning of these values is not essential for the understanding of our model. So we just explain a few:

KKH varies from *cold state*, *heat state*, *above cold below heat*, *above heat below cold*, *real heat false cold*, *real cold false heat*,
to finally *normal*.

[1] In the remainder of this section we closely follow Schnorrenberger [7].

no.	x_i	rule	no.	x_i	rule
1	.80	KKH=hs $\Rightarrow$ KYY=yas	20	.99	KYY=n $\Rightarrow$ YANG=n
2	.80	KKH=ks $\Rightarrow$ KYY=yis	21	.85	KKH=n
3	.80	KOI=os $\Rightarrow$ KYY=yas	22	.85	KOI=n
4	.80	KOI=is $\Rightarrow$ KYY=yis	23	.85	KYY=n
5	.80	KLF=fs $\Rightarrow$ KYY=yas	24	.85	KLF=n
6	.80	KLF=ls $\Rightarrow$ KYY=yis	25	.03	KKH=ohuk $\vee$ KKH=okoh $\vee$ KKH=ekfk $\vee$ KKH=ekfh
7	.99	KKH=hs $\wedge$ KLF=fs $\Rightarrow$ YIN=n	26	.03	KOI=olif $\vee$ KOI=ofil $\vee$ KOI=okih $\vee$ KOI=ohik
8	.99	KKH=hs $\wedge$ KLF=fs $\Rightarrow$ YANG=uem	27	.03	KLF=gm
9	.99	KKH=hs $\wedge$ KLF=ls $\Rightarrow$ YIN=vm	28	.99	KOI=os $\wedge$ KKH=ks $\Rightarrow$ FEVER=lf
10	.99	KKH=hs $\wedge$ KLF=ls $\Rightarrow$ YANG=n	29	.99	KOI=os $\wedge$ KKH=hs $\Rightarrow$ FEVER=sf
11	.99	KKH=ks $\wedge$ KLF=fs $\Rightarrow$ YIN=uem	30	.99	KOI=is $\wedge$ KKH=ks $\Rightarrow$ FEVER=lf
12	.99	KKH=ks $\wedge$ KLF=fs $\Rightarrow$ YANG=n	31	.99	KKH=hs $\wedge$ KLF=fs $\Rightarrow$ FEVER=sf
13	.99	KKH=ks $\wedge$ KLF=ls $\Rightarrow$ YIN=n	32	.99	KKH=ks $\wedge$ KLF=fs $\Rightarrow$ FEVER=lf
14	.99	KKH=ks $\wedge$ KLF=ls $\Rightarrow$ YANG=vm	33	.99	KKH=ks $\wedge$ KLF=ls $\Rightarrow$ FEVER=lf
15	.99	KKH=ehfk $\Rightarrow$ YIN=sm	34	.99	KOI=os $\Rightarrow$ FEVER=lf
16	.99	KKH=ehfk $\Rightarrow$ YANG=se	35	.99	KOI=ofil $\Rightarrow$ FEVER=sf
17	.99	KKH=ehfk $\Rightarrow$ YIN=se	36	.99	KOI=okih $\Rightarrow$ FEVER=mf
18	.99	KKH=ehfk $\Rightarrow$ YANG=sm	37	.99	KKH=ks $\Rightarrow$ FEVER=mf
19	.99	KYY=n $\Rightarrow$ YIN=n			

Figure 2: Rules for Chinese diagnosis

KOI varies from *surface state*, *inside state*, *surface empty inside abundant*, *surface abundant inside empty*, *surface cold inside heat*, *surface heat inside cold*
to finally *normal*.

The remaining symbols have similar explications, yin and yang are measured in a scale from very high to very low and similar is fever, always including the state *normal*.

With these definitions of variables and values the Chinese experts were able to formulate rules which we present in figure 2. The syntax is self-explanatory, the probabilities x_i represent an estimation of the respecting rules to be true in the human population. They are not the result of a statistical analysis and so might involve errors.

A first trial to solve (3) with these rules shows them to be contradictory. IPF does not converge, there is no distribution which represents all rules at the same time. The consulted experts did not coincide in a common estimation of the probabilistic relations among variables and values. This lack gives rise to a revision of the rule probabilities following the process developed in the last section. To do so we choose a partition which isolates the influence of KOI, KKH, KLF over Fever from that of KOI over KYY and KKH, KLF over KYY, YIN, YANG. More precisely the partition is

$$\boldsymbol{U} = \{U_1 = \{28 \text{ to } 37\}, U_2 = \{3,4\}, U_3 = \{8,10,12,14,16,18,20\}, U_4 = \{7,9,11,13,15,17,19\}, U_5 = \{1,2,5,23\}, U_6 = \{21,24,25,27\}, U_7 = \{22,26\}\}.$$

The highest $\bar{x}$ (cf. the last section) for which the parametric problem (6) permits a solution is $\bar{x} = 0.95$. This indicates a low degree of contradiction or vice versa a relative high measure of consistency.

The respective rules including the world variables W_k are then iterated. Evidencing *all* worlds we get the values x_i^* which are compared with x_i in table 3.

Table 3: x_i vs. x_i

rule	3	4	9	15	21	22	24	28	29	32
x_i	.80	.80	.99	.99	.85	.85	.85	.99	.99	.99
x_i^*	.73	.77	.98	1.	.88	.87	.87	.98	.98	1.

Note that only 10 out of 37 rules need to be altered, the rest remains unchanged. If we accept the matching method all experts except those in U_3 and U_5 must agree to corrections. If they do, they now have a knowledge base for Chinese medicine at their disposal.

Let us apply this knowledge base to a concrete patient. He or she shows high fever FEVER = sf and his or her appearance indicates *real heat false cold*. KKH = ehfk. If we inform this evidence to the knowledge base it diagnoses with .87 a severe inner disease KOI = is which should cause further medical examinations.

5. SUMMARY AND FURTHER RESEARCH

In our paper we are concerned with the applicability of a probability distribution as knowledge base in AI. If there is incomplete information about a population's distribution, the arbitrary assessment of missing probabilities might cause undesired dependencies and consequently imply wrong decisions.

These difficulties can be removed by the construction of distributions following the principle of maximum entropy instead of an arbitrary completion of the missing information.

The deduction process of response from such a knowledge base can be likewise performed respecting the ambitious principle of minimum relative entropy.

Both principles, that of maximum entropy and that of minimum relative entropy guarantee the only unbiased knowledge processing, so to say. The expert system shell SPIRIT facilitates such a knowledge processing.

Since the communication language between the entropy-driven expert system shell SPIRIT and the user, allows of the very rich syntax of conditionals on propositional variables this freedom might cause contradictions or inconsistencies.

If contradictory judgements of different experts are the reason for such an inconsistency, it can be solved matching their respective opinions. This matching process is applied to a diagnosis system in Chinese medicine.

To aggregate information shares which in part are contradictory, is an important field of research. The aggregation should always meet ambitious requirements such as a minimal modification of *all* supplied information. Only if a mechanism of that kind is available knowledge of many fields or contexts can be collected and joined free of contradiction. The theoretical development of a suitable aggregation process and the implementation of an expert system shell is an urgent aim of our research.

The reader who is interested in more details on the expert system shell SPIRIT as well as further results in research might contact us via e-mail

wilhelm.roedder@fernuni-hagen.de

or visit our homepage

http://www.fernuni-hagen.de/BWLOR/welcome.htm

REFERENCES

[1] I. Csiszár: *I-Divergence Geometry of Probability Distributions and Minimisation Problems*. The Annals of Probability 3, (1): 146 - 158 (1975).

[2] I. Csiszár: *Why Least Squares and Maximum Entropy? An Axiomatic Approach to Inference for Linear Inverse Problems*. The Annals of Statistics 19 (4): 2032 – 2066 (1991).

[3] G. Kern-Isberner: *Characterising the principle of minimum cross-entropy within a conditional-logical framework*. Artificial Intelligence 98: 169-208 (1998).

[4] S. L. Lauritzen: Graphical Association Models (Draft), Technical Report IR 93-2001, Institute for Electronic Systems, Dept. of Mathematics and Computer Science, Aalborg University (1993).

[5] W. Rödder and G. Kern-Isberner: *Representation and extraction of information by probabilistic logic*. Information Systems 21 (8): 637 – 652 (1996).

[6] W. Rödder and C.-H. Meyer: *Coherent knowledge processing at maximum entropy by SPIRIT*, Proceedings 12th Conference on Uncertainty in Artificial Intelligence, E. Horitz and F. Jensen (editors), Morgan Kaufmann, San Francisco, California: 470 – 476 (1996).

[7] C. C. Schnorrenberger: *Lehrbuch der chinesischen Medizin für westliche Ärzte*, Hippokrates, Stuttgart (1985).

[8] C. E. Shannon: *A mathematical theory of communication*, Bell System Tech. J. 27, 379-423 (part I), 623 – 656 (part II) (1948).

[9] J. E. Shore and R. W. Johnson: *Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross Entropy*. IEEE Trans. Information Theory 26 (1): 26 – 37 (1980).

[10] J. Whittaker: *Graphical Models in Applied Mathematical Multivariate Statistics*, John Wiley & Sons (1990).