

Knowledge Processing under Information Fidelity

Wilhelm Rödder

FernUniversität Gesamthochschule in Hagen
D-58084 Hagen, Germany
wilhelm.roedder@fernuni-hagen.de

Abstract

XSPIRIT is a professional expert system-shell for knowledge acquisition, inference and response using conditional logic and probability. Composed conditionals on propositional variables with finite domain are the communication tool between the user and the knowledge base, making the process of acquisition, inference and query comfortable and intelligible. XSPIRIT allows partial rather than complete information about the knowledge domain and supplements missing parts by the principle of information fidelity. By virtue of evident temporary information, knowledge undergoes a well-defined adaptation process, respecting this principle again. The construction and transformation of probability distributions as developed here, allow acquired knowledge, remaining uncertainty and strength of inference to be measured in the information units [bit]. XSPIRIT allows large-scale applications with hundreds of composed conditionals and umpteen variables.

1 Introduction

Humans memorise and deduce facts »in a conditional environment«. »If A then B « is a constituent of human knowledge. Such knowledge pieces together with concatenation mechanisms build a highly netted memory, the knowledge base. Once situativ things become evident, our knowledge undergoes a complex adaptation process resulting in conclusions. Conclusions cause reactions or response.

Probability distributions on event fields are a suitable means to model such deduction processes in a knowledge base. Bayes nets are the graphical structure which might help to organise knowledge and the corresponding distributions, and make them manageable, cf. [Lauritzen and Spiegelhalter, 1988] or [Pearl, 1988]. Bayes nets allow to break down the global distributions to local marginals, at the cost of conditional independence assumptions which rarely meet reality. Inference in Bayes nets then is the technically complex but unpretentious calculation of conditioned probabilities. We do not follow this idea, but want to enable

the user to formulate any local knowledge pieces without forcing him to construct a directed acyclic graph. After an automatic aggregation of these pieces to a knowledge base he can adapt this knowledge to any situativ modification.

Thus, different from the established form of a probability model, we shall discuss an inference process which

- allows partial rather than complete information about the probability distribution. The missing parts are completed by the principle of information fidelity which turns out to coincide with the principle of minimum relative entropy.
- permits inference from virtual evidence. Vague assumptions about the actual situation can be processed as well as sure things.
- allows the user to »communicate« with the distribution in the comfortable language of composed conditionals.

There is a lot of related work to our paper in literature. To minimize relative entropy when adapting the dependency structure in a distribution to new information was discussed already in early works [Jaynes, 1978], [Cheeseman, 1983], and justified axiomatically [Shore and Johnson, 1980], [Paris and Vencovská, 1990], [Csiszár, 1991], [Kern-Isberner, 1998]. Even the idea to use such adaptation for inference is not new, see [Cheeseman, 1983], [Hájek et al., 1992], [Goldszmidt et al., 1993], [Rödder and Meyer, 1996].

All probabilistic inference presented here, however, is of purely information theoretic nature. The quantity of acquired knowledge can be measured in [bit], as well as the strength of evidence. Knowledge is reduced uncertainty, inference means uncertainty adaptation to new information. And even query and response are information demand and supply.

In section 2 we develop the language of composed conditionals as a communication tool between user and knowledge base, in section 3 we analyse the relations between probability and information, in section 4 we go through the entire process of knowledge acquisition, inference and response, as realized in the expert system shell XSPIRIT. A small example emphasises the conceptual difference between Bayes nets and the inference method developed here. Section 5 gives a short technical introduction to the shell and invites the reader to test it.

2. Conditionals and Probability

2.1 Conditionals in a Three-valued Logic

Concerning the semantics we consider a population of individuals or objects with their respective properties. Our intention is the description of relations between these properties. As far as the syntax concerns we model the knowledge domain as a finite set of finite-valued variables $V = \{V_1, \dots, V_L\}$ and the respective event algebra F on the elementary events $v = v_1 \dots v_L$, here v_l are values of V_l . As usual Ω is the all- and $\emptyset$ the empty event.

Each event from F can be identified in a natural way with a propositional sentence, built from literals $V_l = v_l$ by the connectives negation, conjunction and disjunction, cf. [Rödder and Kern-Isberner, 1997]. Simple conjuncts of such literals we often write as unordered tuples of the respective values, $v_1 \dots v_L$, for instance. Hence each event may be expressed in an intelligible way, which we shall demonstrate by example 1 below. We do not distinguish between events and propositional sentences, further on.

Generic events from F or the corresponding propositions we denote as capital letters, indexed if necessary: $A, B, \dots$, $A_i, B_i, \dots$, E_i, F_i . If $v \subset A$, we call the proposition A true in the world v and false, otherwise: $A(v) = t$ or $A(v) = f$. If A is true (false) on the entire Ω , we write $A = t$ ($A = f$). The set of all propositional sentences forms the language L .

Calabrese [Calabrese, 1991] and earlier de Finetti [deFinetti, 1972] go a step further and define conditionals $B|A$ for B and A from L . Calabrese in [Calabrese, 1991] and [Calabrese, 1994] derives a rich language of composed conditionals. We go through that in more detail now.

The exact definition of a conditional reads

$$B|A(v) = \begin{cases} t & \text{if } v \subset AB \quad (= A \wedge B) \\ f & \text{if } v \subset A\bar{B} \quad (= A \wedge \bar{B}) \\ u & \text{if } v \subset \bar{A}. \end{cases} \quad (1)$$

Barring means negation and t(true), f(false), u(undefined) are values in a three-valued logic, cf. [Rescher, 1969]. In this logic the connectives $\neg, \wedge, \vee$ apply for t, f as usual and for u as follows:

$$t \wedge u = t \vee u = t, \quad f \wedge u = f \vee u = f, \quad u \wedge u = u \vee u = u, \quad \bar{u} = u. \quad (2)$$

The truth-value *undefined* does not change neither t, f nor u under conjunction or disjunction.

The set of all conditionals $B|A$ is the language $L|L$. Identification of $B|\Omega$ with B embeds L in $L|L$. Similar to above we write $B|A = t|A$ ($= f|A$) if the conditional is true (false) on the whole A .

Following Calabrese further, we compose conditionals by $\neg, \wedge, \vee$ and $|$ to get a very rich linguistic structure for the communication between the user and the knowledge base, as shall be developed in the remainder of this section.

With the conventions in Reschers three-valued logic (2), the author in [Calabrese, 1991] defines pointwise:

$$\bullet \quad [(B|A) \wedge (D|C)](v) = (B|A)(v) \wedge (D|C)(v),$$

$$\bullet \quad [(B|A) \vee (D|C)](v) = (B|A)(v) \vee (D|C)(v), \quad (3)$$

$$\bullet \quad (\overline{B|A})(v) = \overline{(B|A)(v)}.$$

By these pointwise attributions any conjunct, disjunct or negation of conditionals is well-defined also globally.

Let the conditional operator $|$ apply as in the following scheme:

$ $	t	f	u
t	t	u	t
f	f	u	f
u	u	u	u

(read e.g. $f|u = f$). (4)

With this convention a pointwise definition of a conditioned conditional reads:

$$\bullet \quad [(B|A)|(D|C)](v) = (B|A)(v) | (D|C)(v). \quad (5)$$

Pointwise attribution of truth-values again permits the global definition of conditioned conditionals on Ω . More on the hierarchy of composed conditionals the reader might find in Calabrese's original paper, as well as the proof of the following theorem 1. Theorem 1 permits the reduction of composed conditionals to simple ones.

Theorem 1 (Composed conditionals)

With the conventions in (2), (4) and the definitions in (3), (5) the following equations hold:

$$\begin{aligned} \bullet \quad & (B|A) \wedge (D|C) = [(B \vee \bar{A})(D \vee \bar{C})] | A \vee C, \\ \bullet \quad & (B|A) \vee (D|C) = (AB \vee CD) | A \vee C, \\ \bullet \quad & (B|A) | (D|C) = B|A(D \vee \bar{C}). \end{aligned}$$

Composed conditionals are reducible to simple ones, but in their composed form allow an immense linguistic richness, as demonstrates the following example.

Example 1 (Linguistics and composed conditionals)

With the use of mnemonic letters the following propositions should be self explicatory:

$$\begin{aligned} \bullet \quad & [(M|(C_1 \wedge C_2)) \wedge ((C_1 \wedge C_2)|M)](v), \\ & \text{in the world } v \text{ machine } M \text{ is o.k. iff the component } C_1 \\ & \text{is o.k. and } C_2 \text{ is o.k.} \\ \bullet \quad & [((G \vee A) \wedge (B|G)) | S](v), \\ & \text{in the world } v \text{ Herr Schmid is German or Austrian, if} \\ & \text{he is German he comes from Berlin.} \\ \bullet \quad & [(C \vee (V | \bar{C})) | L](v), \\ & \text{in the world } v \text{ Mr. Lee is Chinese, if not, he is} \\ & \text{Vietnamese.} \end{aligned}$$

The reader might notice that we can make global considerations rather than »in the world v «, too. In a certain population or knowledge context, the machine M is o.k. iff C_1 and C_2 are o.k. with a probability of 90%, say. Probabilities of conditionals are constituents of our AI-concept. More on that in the next sections, cf. also [Rödder, 2000]. Because of theorem 1 it suffices to consider simple conditionals of the form $B|A$ further on. Bear in mind, however, that they represent the total hierarchy of arbitrarily composed conditionals and their respective linguistic pendants.

In what follows we define sets of conditionals with certain properties. These properties make them »a conditioning partition« on Ω .

Definition 1 (c-independent and disjoint conditionals)

- i.) $B|A$ is conditionally or c-independent of $D|C$ iff $(B|A)(D|C) = B|A$, and c-dependent otherwise.
- ii.) $B|A$ and $D|C$ are disjoint iff $ABCD = f$.

Definition 2 (Conditional basic set CBS)

- i.) A set S of conditionals is sufficient for Ω , if any $v \subset \Omega$ can be expressed as a conjunct of elements from S .
- ii.) A conditional basic set CBS is a set of conditionals sufficient for Ω such that, for any $B|A, D|C \in \text{CBS}$ one of three statements holds:
 $B|A$ and $D|C$ are disjoint,
 $B|A$ is c-independent of $D|C$,
 $D|C$ is c-independent of $B|A$.

The following example gives an idea of conditional basic sets.

Example 2 (Conditional basic sets)

- i.) $\{v = v_1 \dots v_L\}$ is a CBS.
- ii.) $\{v_1\} \cup \{v_2|v_1\} \cup \dots \cup \{v_L|v_{L-1} \dots v_1\}$ is a CBS. Note that conditionals "more to the right" are c-independent of those "more to the left", if they include no distinct values $v_i \neq v_i'$. If they do include distinct values, they are disjoint. Any v can be resolved into "factors" $v = v_1 \wedge (v_2|v_1) \wedge \dots \wedge (v_L|v_{L-1} \dots v_1)$. To show this, use (1) and (2).
- iii.) $\{v_1 \dots v_l\} \cup \{v_{L-l+1} \dots v_L | v_1 \dots v_l\}$ is a CBS, as shows a similar reasoning like that in ii).
- iv.) ii) or iii) for an arbitrary permutation $l_1 \dots l_L$ of $1 \dots L$ is a CBS.

Conditional basic sets will serve as a means for measuring the intrinsic uncertainty in a probability distribution on Ω and this measure will turn out to be independent of the special choice of the CBS.

2.2 The Probability of Conditionals

A probability measure P on Ω can be generalized to conditionals by the natural requirement, that for the conjunct of c-independent conditionals the probabilities should multiply. As $B|A$ is c-independent of A and because $(B|A) \wedge A = BA$, it follows immediately $P(BA) = P((B|A) \wedge A) = P(B|A) \cdot P(A)$ and hence $P(B|A) = P(BA) / P(A)$, if only $P(A) > 0$. The probability of a conditional is its conditioned probability.

From the logical factorisation

$$v = v_1 \wedge (v_2|v_1) \wedge \dots \wedge (v_L|v_{L-1} \dots v_1) \quad (\text{cf. example 2})$$

we get the well-known probabilistic factorisation

$$P(v) = P(v_1) \cdot P(v_2|v_1) \cdot \dots \cdot P(v_L|v_{L-1} \dots v_1).$$

Probabilistic independence reflects the logical-based c-independence, as it should be.

3. Probability, Uncertainty and Information

3.1 Basic Considerations

Entropy first 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. When the engineer Shannon [Shannon, 1948] founded information theory, he looked for an adequate denotation of $H(P) = -\sum_v P(v) \cdot \text{ld } P(v)$, where ld is the dual logarithm and $0 \cdot \text{ld } 0 = 0$. In this theory signals v vary in a finite alphabet $v \in V$ and H measures the average information rate which can be transmitted from a source via a channel to a receiver. John v. Neumann suggested to name it *entropy*, following an anecdote in the Scientific American [Horgan, 1990] because "Shannon would have an advantage in debates about this theory."

In the remainder of this section we shall study entropy further.

If we learn that a proposition A with probability $P(A)$ becomes true, we receive the information $-\text{ld } P(A)$ [bit]. Information is reduction of uncertainty, any textbook of information theory goes through a proof of this fact. We learn that a conditioned proposition $B|A$ becomes true if BA becomes true, cf. (1). Though the reduced uncertainty is not $-\text{ld } P(BA)$ but rather $-\text{ld } P(BA) - (-\text{ld } P(A))$. This result comes from the additivity of uncertainty reduction for c-independent conditionals. We put it as a lemma and leave the obvious proof to the reader.

Lemma 1 (Information of c-independent conditionals) If for any c-independent $B|A$ of $D|C$ or $D|C$ of $B|A$, the uncertainty reduction $\text{inf } (B|A \wedge D|C)$ equals $\text{inf } (B|A) + \text{inf } (D|C)$, it follows $\text{inf } (B|A) = -\text{ld } P(BA) - (-\text{ld } P(A))$.

Let us now consider a similar situation as before, but – for whatever reason – our situativ conviction of $B|A$ to become true, has changed from $P(B|A)$ to $Q(B|A)$. This change is change of uncertainty and the difference equals $\Delta \text{inf} = -\text{ld } P(B|A) - (-\text{ld } Q(B|A)) = \text{ld } (Q(B|A) / P(B|A))$, cf. figure 1.

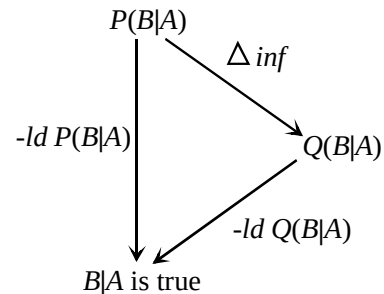


Figure 1: Sequential information gain

Note that the expectation of this uncertainty change is $Q(BA) \cdot \text{ld } (Q(B|A) / P(B|A))$, as $Q(BA)$ is the probability of $B|A$ to become true in the light of our new conviction.

3.2 Information Adaptation

In this section we study the adaptation of a distribution P to new information, supplied in the form of conditionals from $\mathcal{L}|\mathcal{L}$ and their respective desired probabilities. To distinguish between such conditionals and the elements of a CBS, we use lowercase letters $f|e$ for the latter. There should be no problems with this notation.

The following theorem shows that the overall expected uncertainty change does not depend on a specific CBS.

Theorem 2 (Unique uncertainty change)

Let CBS and CBS' be two arbitrary conditional basic sets with generic conditionals $f|e$ and $f'|e'$, respectively. Let Q and P be two distributions on Ω , such that Q is totally continuous with respect to P . Then the following equality holds ($0 \cdot \text{ld}(0/0) = 0$):

$$\sum_{f|e \in \text{CBS}} Q(f|e) \cdot \text{ld}(Q(f|e)/P(f|e)) = \sum_{f'|e' \in \text{CBS}'} Q(f'|e') \cdot \text{ld}(Q(f'|e')/P(f'|e')).$$

Proof:

$$\begin{aligned} & \sum_{f|e \in \text{CBS}} Q(f|e) \cdot \text{ld}(Q(f|e)/P(f|e)) = \\ & \sum_{f|e \in \text{CBS}} \left(\sum_{v \subset f|e} Q(v) \right) \cdot \text{ld}(Q(f|e)/P(f|e)) = \\ & \sum_v Q(v) \cdot \left(\sum_{f|e \text{ with } v \subset f|e} \text{ld}(Q(f|e)/P(f|e)) \right) \end{aligned}$$

For each v the inner summation runs through conditionals, which are not disjoint and hence pairwise c -independent (one independent of the other). Because a CBS is sufficient we even have

$$v = \bigwedge_{v \subset f|e} f|e.$$

As furthermore logical independence implies probabilistic independence we get

$$\sum_v Q(v) \cdot \text{ld} \prod_{f|e \text{ with } v \subset f|e} (Q(f|e)/P(f|e))$$

$$\text{and finally } \sum_{v \in \Omega} Q(v) \cdot \text{ld}(Q(v)/P(v)).$$

The CBS was chosen arbitrarily, which completes the proof.

The basic idea of the proof was to show that either side of the equation equals $\sum_{v \in \Omega} Q(v) \cdot \text{ld}(Q(v)/P(v))$.

This expression is known as relative entropy $R(Q, P)$ between P and Q or directed divergence from P to Q . Expected uncertainty change from P to Q in any CBS equals the relative entropy between P and Q .

If $P = P^0$ is the uniform distribution, $H(Q) = \text{ld} |\Omega| - R(Q, P^0)$ is the (absolute) entropy of Q . Thus the uncertainty

change from P^0 to Q is also $\text{ld} |\Omega| - H(Q)$. Both R and H measure in [bit].

As mentioned at the beginning of the present section we now study the problem of adapting the distribution P to new information supplied as a set of conditionals with desired probabilities $x_i \in [0, 1]$: $\mathfrak{R} = \{B_i | A_i[x_i]\}$.

We demand this adaptation to be such that the former uncertainty structure in P is preserved as far as possible, i.e. we solve

$$\bar{P} = \arg \min R(Q, P) \text{ s.t. } Q(B_i | A_i) = x_i, i = 1 \dots I, \quad (6)$$

and call $\bar{P}$ the $\mathfrak{R}$ -adaptation of P . As (6) preserves the expected uncertainty reduction for any CBS as far as possible, it obeys the *principle of information fidelity*. There are axiomatic accesses to conditional knowledge adaptation, cf. [Csisár, 1991], [Shore and Johnson, 1980], [Paris and Vencovská, 1990], [Kern-Isberner, 1998]. The objective of our approach is to focus upon a cautious information work up. With the presentation of all logical, probabilistic and information-theoretic prerequisites we are now ready to develop knowledge processing as realized in XSPIRIT.

4 Knowledge Acquisition, Inference and Response

Probabilistic knowledge processing in XSPIRIT is performed in the four steps: Initialisation, knowledge acquisition, inference and response. We shall go through these steps in detail now.

Initialisation

Here we commit ourselves to a knowledge domain by defining the variables V_i and their respective values v_i . This determines $\Omega = \{v\}$ and the event algebra $\mathcal{F}$. The uniform P^0 on $\mathcal{F}$ means absolute ignorance on $\mathcal{F}$, as the variables V_i are mutually independent from each other.

Knowledge Acquisition

The knowledge engineer supplies (composed) conditionals and their respective desired probabilities $\mathfrak{R} = \{B_i | A_i[x_i], i = 1 \dots I\}$. $B_i | A_i[x_i], i = 1 \dots I$ is the imperative to adapt P^0 to $\mathfrak{R}$. As developed in the last section we solve

$$P^* = \arg \min R(Q, P^0) \text{ s.t. } Q(B_i | A_i) = x_i, i = 1 \dots I. \quad (7)$$

P^* is the adapted knowledge following the principle of information fidelity. $R(P^*, P^0)$ measures the knowledge increment in [bit]. Remind that minimizing relative entropy is equivalent to maximizing entropy, and verify the equation $H(P^0) - R(P^*, P^0) = H(P^*)$. Actual uncertainty is maximum uncertainty minus acquired knowledge.

Inference

The user supplies (composed) conditionals and their respective temporarily evident probabilities $\mathfrak{E} = \{F_j | E_j[y_j], j = 1 \dots J\}$. $F_j | E_j[y_j], j = 1 \dots J$ is the imperative to find a probability distribution which meets evidence. Evidence is not additional knowledge to be acquired, but describes a

temporary reflection about things that might happen. "If very likely with $y = 90\%$ a person's nationality were German, which conclusions could we infer from that?" Evident conditionals and their probabilities in almost all cases are contradictory to the knowledge supplied by $\mathfrak{K}$! Of course in our knowledge domain there is no 90% probability to be German, in general. Because of the observations in the previous section we calculate

$$P^{**} = \arg \min R(Q, P^*) \text{ s.t. } Q(F_j | E_j) = y_j, j = 1 \dots J. \quad (8)$$

P^{**} has minimal directed divergence from P^* , subject to evidence. Among all Q , P^{**} is the distribution with the minor change of expected uncertainty relative to P^* . $R(P^{**}, P^*)$ is the strength which evidence puts upon P^* to infer P^{**} , and measures in [bit].

Response

The user wants, in the evident situation, the knowledge base to answer a question, which also might be of the conditional form $(H|G)$. The response is $P^{**}(H|G)$. Response is inferred from the knowledge base $\mathfrak{K}$ and from evidence $\mathfrak{E}$.

While (7) and (8) is inference, as (former) knowledge is adapted to new information in a sophisticated transformation process, response is not. Response is the calculation of a conditional probability and hence the mere evaluation of a distribution.

W. Rödder and G. Kern-Isberner in [Rödder and Kern-Isberner, 1997] discuss the example Lea Sombé of people in a fictitious society, which originally was presented in [Sombé, 1992].

Example 3 (Lea Sombé)

In a fictitious society we consider the properties: to be a student or not $S = s/\bar{s}$; to be young or not $Y = y/\bar{y}$; to have the marital status single, married or corporate life $M = s, m, c$; to be parent or not $P = p/\bar{p}$. The following information is given:

- about 90% of all students are young,
- about 80% of all young people are single,
- 70% of singles are young,
- 30% of young people study,
- members of the society which live in corporate lives, are young with 80% ,
- students with children, in 90% of all cases are married or live in corporate lives.

In the notation of the present paper we have the conditionals:

$$Y = y|S = s[.9], \quad M = s|Y = y[.8], \quad Y = y|M = s[.7], \\ S = s|Y = y[.3], \quad Y = y|M = c[.8], \quad \overline{M = s}|S = s \wedge P = p[.9].$$

After solving (7) to determine P^* , we want to make conclusions about Lea Sombé who evidently is a student and has a child. Is she young? Calculating P^{**} in (8) under

certain evidence and evaluating $P^{**}(Y=y)$ yields 81% . Lea Sombé very likely is young.

The information in P^* only counts 1.05 [bit], the remaining uncertainty about the fictitious society is high, namely 3.53 [bit]. Imposing evidence on P^* results in an information flow of 1.62 [bit]. All results were calculated with the shell XSPIRIT.

Besides of its simplicity example 3 is not accessible for Bayes nets, for two reasons:

- information about the society is incomplete inasmuch as it does not fully determine a probability distribution,
- supplied conditioned probabilities do not build a directed acyclic graph.

You can handle virtual evidence and augment the number of variables and conditionals almost arbitrarily. XSPIRIT is able to treat hundreds of composed conditionals on umpteen variables, as we shall relate in the next section.

5 The Shell XSPIRIT

The shell XSPIRIT 3.0, a JAVA-version, is a professional tool for knowledge processing under information fidelity. To acquire knowledge the user defines variables and supplies conditionals in a comfortable syntax. The syntax permits conjunctions, disjunctions and negations of literals and their respective composed conditionals, in an arbitrary hierarchy. Variables might be boolean, nominal and ordinal. The attributes might be associated with real numbers (utilities) allowing the calculation of expected utility for respective decision models. The shell supports an interactive process to untie inconsistent information if provided by the user. For an application to on-line banking and credit worthiness confer [Kulmann and Reucher, 2000], for a business-to-business-approach see [Kulmann and Rödder, 2001].

Once all conditionals and their respective probabilities are supplied to the system, XSPIRIT automatically generates a suitable graphical structure, a hypergraph, and builds the knowledge base P^* by virtue of corresponding marginal distributions on the hyperedges –sometimes called LEGs (local events groups). In the presence of evidence XSPIRIT supplies P^{**} and evaluates any query, asked by the user. All communication is performed in the rich syntax of composed conditionals. P^* and P^{**} are calculated by a generalised Iterative Proportional Scaling (IPS) procedure, whose convergence was proved by Csiszár [Csiszár, 1975]. The system at any time informs about the amount of information already acquired and the remaining uncertainty in [bit]. For more mathematical details see [Rödder and Meyer, 1996].

Bayes nets are importable and XSPIRIT also provides a frequentistic learning process. If knowledge about the domain under consideration is available as a sample of elementary events rather than estimated probabilities x_i , the shell adopts this information in a separate module.

The reader interested in the shell XSPIRIT might visit the homepage

<http://www.xspirit.de>.

References

- [Calabrese, 1991] P. G. Calabrese: *Deduction and Inference Using Conditional Logic and Probability*, in: Conditional Logic in Expert Systems, I. R. Goodman, M. M. Gupta, H. T. Nguyen and G. S. Rogers (editors). Elsevier Science Publishers B. V., 71 – 100, (1991).
- [Calabrese, 1994] P. G. Calabrese: *Theory of Conditional Information with Applications*, in: Special Issue on Conditional Event Algebra, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 24, No. 12, 1676 – 1684, (1994).
- [Cheeseman, 1983] P. Cheeseman: *A method of computing generalized Bayesian probability values for expert systems*, Proc. 6th Int. Joint Conf. on AI (IJCAI – 83), Karlsruhe, (1983).
- [Csiszár, 1975] I. Csiszár: *I-divergence Geometry of Probability Distribution and Minimisation Problems*. The Annals of Probability, 3, No.1, 146 – 158, (1975).
- [Csiszár, 1991] 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).
- [deFinetti, 1972] B. de Finetti: *Induction and Statistics*, Wiley, New York, (1972).
- [Goldszmidt et al., 1993] M. Goldszmidt, P. Morris, J. Pearl: *A Maximum Entropy Approach to Nonmonotonic Reasoning*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 15, No. 3, 220-232, (1993).
- [Hájek et al., 1992] P. Hájek, T. Havránek, R. Jirousek: *Uncertain Information Processing in Expert Systems*, CRC-Press, (1992).
- [Horgan, 1990] J. Horgan: *Claude E. Shannon, Unicyclist, juggler and father of information theory*, PROFILE in: Scientific American, 16 – 17, (1990).
- [Jaynes, 1978] E.T. Jaynes: *Where do stand on maximum entropy?*, in: The Maximum Entropy Formalism, R.D. Levine and M. Tribus (editors), Cambridge Mass., MIT-Press, (1978).
- [Kern-Isberner, 1998] G. Kern-Isberner: *Characterising the principle of minimum cross-entropy within a conditional-logical framework*, Artificial Intelligence, Vol. 98, 169 – 208, (1998).
- [Kulmann and Rödder, 2001] F. Kulmann and W. Rödder: *Probabilistische Modellbildung auf der Basis von Scoring-Schemata*, Proc. GOR Operations Research, Dresden, 477-482, (2001).
- [Kulmann and Reucher, 2000] F. Kulmann and E. Reucher: *Computergestützte Bonitätsprüfung bei Banken und Handel*, Die Betriebswirtschaft (DBW), 60, 113-122, (2000).
- [Lauritzen and Spiegelhalter, 1988] S.L. Lauritzen and D. J. Spiegelhalter: *Local computations with probabilities in graphical structures and their applications to expert systems*, Journal of the Royal Statistical Society, 13 (2), 415 – 448, (1988).
- [Pearl, 1988] J. Pearl: *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, San Mateo Cal., (1988).
- [Paris and Vencovská, 1990] J.B. Paris, A. Vencovská: *A note on the inevitability of maximum entropy*, Int. J. of Approximate Reasoning, 14, 183-223, (1990).
- [Rescher, 1969] N. Rescher: *Many-Valued Logics*, Mc Craw-Hill, New York, (1969).
- [Rödder, 2000] W. Rödder: *Conditional Logic and the Principle of Entropy*, Artificial Intelligence, 117, 83 – 106, (2000).
- [Rödder and Kern-Isberner, 1997] W. Rödder and G. Kern-Isberner: *Léa Sombé und entropie-optimale Informationsverarbeitung mit der Expertensystem-Shell SPIRIT*, OR Spektrum, 19, 41 – 46, (1997).
- [Rödder and Meyer, 1996] W. Rödder and C.-H. Meyer: *Coherent knowledge processing at maximum entropy by SPIRIT*, Proceedings 12th Conference on Uncertainty in Artificial Intelligence, E. Horvitz and F. Jensen (editors), Morgan Kaufmann, San Francisco Cal., 470 – 476, (1996).
- [Shannon, 1948] C. E. Shannon: *A mathematical theory of communication*, Bell System Tech. J., 27, 379 – 423 (part I), 623 – 656 (part II), (1948).
- [Shore and Johnson, 1980] 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).
- [Sombé, 1992] L. Sombé: *Schließen bei unsicherem Wissen in der Künstlichen Intelligenz*, Vieweg, Braunschweig, Wiesbaden, (1992).