WHERE WE STAND AT PROBABILISTIC REASONING

Wilhelm Rödder, Elmar Reucher and Friedhelm Kulmann
Department of Operations Research, University of Hagen, Profilstr. 8, 58084 Hagen, Germany

Keywords: Probabilistic Reasoning, Bayes-Nets, Entropy, MinREnt-Inference, Expert-System SPIRIT.

Abstract: Bayes-Nets are a suitable means for probabilistic inference. Such nets are very restricted concerning the communication language with the user, however. MinREnt-inference in a conditional environment is a powerful counterpart to this concept. Here conditional expressions of high complexity instead of mere potential tables in a directed acyclic graph, permit rich communication between system and user. This is true as well for knowledge acquisition as for query and response. For any such step of probabilistic reasoning, processed information is measurable in the information theoretical unit [bit]. The expert-system-shell SPIRIT is a professional tool for such inference and allows realworld (decision-)models with umpteen variables and hundreds of rules.

1 THINKING AND EXPERT SYSTEMS

1.1 From the Human Expert to his Artificial Counterpart

Humans' capabilities to memorize and recall knowledge and images, to infer facts from other facts, and to justify or explain their conclusions are admirable. The most surprising is man's ability concerning nonmonotonic reasoning: An ostrich is a bird and "all" birds fly but an ostrich does not, is contradictory but nevertheless accepted even by little childs (Rödder and Kern-Isberner, 2003a), p. 385. It was a long and a painful way for scientists to understand all such capabilities and to do first steps in the direction of modeling them. Respective studies fructified significantly artificial intelligence in its effort to simulate such phenomena on the computer. From this research resulted a great number of computer programs, called expert-systems.

1.2 Milestones in the History of Expert-Systems

After the overwhelming enthusiasm in the scientific community after the 1956 AI-workshop in Dartmouth, very famous researchers in the AI-field experimented with expert-systems like Advice Taker (1958) by McCarthy, General Problem Solver (1960s) by Newell and Simon, Mycin (1972) by Buchanan and Shortcliff, Prosepector (1979) by Duda. Further projects were Dendral, Drilling Advisor etc. For a more extensive discussion see (Harmon and King, 1985). Duda proposed a modified bayesian concept to calculate the strength of rules in a rule based system. As Duda's concept often did not show up comprehensible results a new generation of probabilistic expert-systems came up.

Scientists tried to beat the difficulties of modeling human thinking by various concepts: propositional logics, predicate logics, default logics, circumscription, conditional logics, uncertainty logics, rough sets, trues maintenance systems, among others. A still actual overview of such concepts the reader might find in (Sombe, 1990), even if already published in 1990. Only very few ideas, however, resulted in computer programs able to handle large scale knowledge domains and at the same time simulate human thinking in an adequate way.

It was in the late 1980s and in the 1990s that purely probabilistic concepts for expert-systems have been developed: HUGIN since 1989 (Hugin, 2009) and SPIRIT since 1997 (Spirit, 2009). Even if both expert systems permit probabilistic reasoning they follow absolute different philosophies, however.

Rödder W., Reucher E. and Kulmann F. (2009).
WHERE WE STAND AT PROBABILISTIC REASONING.
In *Proceedings of the 6th International Conference on Informatics in Control, Automation and Robotics - Intelligent Control Systems and Optimization*, pages 394-397
DOI: 10.5220/0002241503940397

2 PROBABILISTIC REASONING

2.1 Probabilistic Reasoning in Bayes-Nets

Following Jensen (Jensen, 2002), p. 19, a bayesian network is characterized as follows:

- a set of finite valued variables linked by directed edges,
- the variables and the edges form a directed acycled graph,
- to each variable with its parents there is attached a potential table,
- the variables might be of type decision variable, utility variable or state variable.

For a deeper discussion of traditional Bayes-Nets confer (Jensen, 2002). Such nets can be formed by an expert as well by empirical data. Later versions of the expert-system shell HUGIN also permit continues rather than discrete variables, only. The great advantage of such a Bayes-Net is the stringent (in)dependency-structure. This is advantageous, in as much as it forces the user to a likewise strict modeling of reality. The advantage might turn into a disadvantage when the user does not dispose of all desired probabilities. Such a model of reality feigns an epistemic state about the knowledge domain which is a biased image of reality, and consequently causes erroneous results when predicting facts from evident facts. Rödder and Kern-Isberner (Rödder and Kern-Isberner, 2003a), p.385, formulate ”Inference is more. Inference is the result of the presumption and logical entailment about the vague population of our perception or even contemplation. Inference takes place in spite of incomplete information about this population.” Probabilistic reasoning under maximum entropy and/or minimum relative entropy, respectively, is a promising alternative to overcome this flaw.

2.2 Probabilistic Reasoning under MinREnt

To build a knowledge base it needs a finite set of finite valued variables $\mathbf{V} = \{V_1, ..., V_L\}$ with respective values v_l of V_l. The variables might be of type type boolean, nominal or numerical. From literals of the form $V_l = v_l$, propositions $A, B, C, ...$ are formed by the junctors $\wedge$ (and), $\vee$ (or), $\neg$ (not) and by respective parentheses. Conjuncts of literals such as $\mathbf{v} = v_1, ..., v_L$ are elementary propositions, $\mathbf{V}$ is the set of all $\mathbf{v}$. $|$ is the directed conditional operator; formulas as $B|A$ are conditionals. The degree to which such conditionals are true in the knowledge domain might be expressed by probabilities $x \in [0;1]$; such conditionals or facts we write $B|A\,[x]$. As to the semantics a model is a probability distribution P for which such conditional information is valid.

More precisely, probabilistic reasoning under MinREnt is realized as follows (Rödder et al., 2006):

1. Definition of the knowledge domain
   Specification of the variables V_l and their respective values v_l, providing the set of all complete conjuncts $\mathbf{v}$.

2. Knowledge Acquisition
   Knowledge acquisition bases on a set of conditionals or facts $\mathsf{R} = \{B_i|A_i\,[x_i], i = 1, \ldots, \mathrm{I}\}$, provided by the user. The solution

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

   is an epistemic state among all Q which minimizes the relative entropy or Kullback-Leibler-divergence R from P^0, satisfying the restrictions R. P^* obeys the MinREnt-principle, in that it respects R without adding any unnecessary information (Rödder and Kern-Isberner, 2003b), p. 467. Bear in mind that for a uniform P^0, minimizing the relative entropy (1) is equivalent to maximizing the entropy $H(Q)$. For more details about the principles MinREnt and MaxEnt and their axiomatic foundations, the reader is referred to (Kern-Isberner, 1998), (Shore and Johnson, 1980).

3. Query
   The query process consists of three steps: focus, adaptation to the focus and question plus response. A focus specifies a temporary assumption about the domain represented by a set of conditionals or facts $\mathsf{E} = \{F_j|E_j\,[y_j], j = 1, \ldots, \mathrm{J}\}$. The adaptation of P^* to E yields the solution

   $$P^{**} = \arg\min R(Q, P^*), \quad s.t.\, Q \models \mathsf{E}. \qquad (2)$$

   Finally for a question $H|G$ under the facts R and the focus E, the answer is

   $$P^{**}(H|G) = z. \qquad (3)$$

The three-step process (1), (2), (3) is called MinREnt inference process (Rödder and Kern-Isberner, 2003b), p. 467. All values of the objective functions in the three steps −as well as the entropies $H(P^0)$, $H(P^*)$, $H(P^{**})$− measure in [bit]; the lower entropy the richer acquired knowledge about the domain. This proximaty to information theory is essential but can not be developed here. For an extensive discussion cf.(Rödder and Kern-Isberner, 2003a).

3 WHAT IS THE ADVANTAGE OF MinREnt OVER BAYES-NETS?

In this section we want to justify our position that MinREnt is better than Bayes-Nets.

- Already the propositions A, B and the conditionals $B|A$ may be pretty complex, due to arbitrary combinations of literals by the conjuncts $\wedge, \vee, \neg$. Handling such expressions in Bayes-Nets is impossible.
- Moreover, syntactical formulas like$(B|A) \wedge (D|C)$, $(B|A) \vee (D|C)$, $\neg(B|A)$, $(B|A)|(D|C)$, so called composed conditionals, allow a rich linguistic semantics on a domain, near human language. For a deeper discussion , cf. (Rödder and Kern-Isberner, 2003a), p. 387. Are already neither general propositions nor conditionals representable in Bayes-Nets the less are composed conditionals.
- The formulation of cyclic dependencies between propositions, e.g. $B|A, C|B, A|C$ is possible in MinREnt-inference. Such dependencies are not permitted in Bayes-Nets as they are DAGs.
- Bayes-Nets suffer from certain difficulties when there is a multiple functional dependence between input variables and an output variable. Such a situation forces the user to additional constructions like ”noisy-and” or ”noisy-or” (Diez and Galan, 2003). In a MinREnt and conditional environment such dependencies are simply and solely formulated as conditionals and the rest is done by the entropy.

All such advantages over Bayes-Nets, of course, must be accompanied by some disadvantages. Because of the absolute freedom in formulating rules, for the unexperienced user there is a high risk to cause inconsistencies: Equation (1) is not solvable. To overcome this problem, SPIRIT allows for solving the inconsistency problem in that it offers slightly modified probabilites $x_i^{'}$ instead of x_i for (1). And the user might decide if he or she accepts these probabilities or not. Usually a set R of rules does not fully determine the epistemic state P over a domain. The freedom to admit imperfect information in R has its price. This price is a possible unreliability of the answer (3). SPIRIT informs the user about such unreliability or second order uncertainty, and invites him/her to add further information.

4 PROFESSIONALITY OF SPIRIT

SPIRIT is a professional expert-system-shell, allowing for the implementation of middle and large scaled knowledge bases. For the reader familiar with probabilistic inference models, first designed for Bayes-Nets, (Hugin, 2009), we list a few examples which where adapted to SPIRIT. Note that the stringent syntax in Bayes-Nets is overcome in SPIRIT. But vice versa, any Bayes-Net application can be modeled in the shell. The models blue baby (BB), troubleshooter (TS), and car repair (CR) are well known, (Breese and Heckerman, 1996), (Hugin, 2009). There are two models in which utility and decision variables are explicitly involved, namely the well known oil drilling problem (OD) and a credit worthiness support system (CW)(Raiffa, 1990). Besides all well known applications an outstanding knowledge base of a business-to-business approach (BS)) was modeled in SPIRIT. The latter with 86 variables and 1051 rules, partly cyclic. Knowledge acquisition for all the models counted in milliseconds (Rödder et al., 2006). All models are available at (Spirit, 2009) and can be tested by the reader. In Table 1 we provide a few data concerning these models. For models with up to umpteen variables and hundreds of rules a suitable form of user interface is necessary so as to inform about the knowledge structure and the inference process.

Table 1: Data for middle and large-scale models, implemented in SPIRIT.

Model	no. variables	no. rules	no. LEGs	$H(P^0)$ [bit]	$H(P^*)$ [bit]
BB	20	340	17	29.91	18.57
TS	76	574	50	76.00	12.83
CR	18	38	13	22.68	6.00
BS	86	1051	36	104.79	87.12
OD	6	18	3	8.17	4.08
CW	10	31	6	11.00	7.38

For this purpose the shell SPIRIT disposes of various communication tools: A list of all variables and their attributes, a list of all conditionals provided by the user, a dependency graph showing the Markov-Net of all stochastic dependencies between such variables, the junction-tree of variable clusters −so called Local Event Groups LEGs− indicating the factorization of the global by marginal distributions, among others (Rödder et al., 2006).

5 CONCLUSIONS AND THE ROAD AHEAD

Knowledge processing in a conditional and probabilistic environment under maximum entropy and minimum relative entropy, respectively, is a powerful instrument supporting the user in various economical and technical decision situations. SPIRIT is a comfortable and professional shell for such knowledge processing.

Recent developments in the field are in test-stage, such as:

- an adaptation of the system to permit simulations of cognitive processes, (Rödder and Kulmann, 2006),
- the calculation of the transinformation or synentropy between arbitrary groups of variables.

Actual research activities are:

- the removal of an eventual unreliability of answers by the initiation of a self-learning process. The theoretical basis for this concept was published already in 2003 (Rödder and Kern-Isberner, 2003b), the implementation in SPIRIT is in the pipeline,
- handling of a mixture of continuous and discrete variables (Singer, 2008).

With such features the expert-system-shell SPIRIT will become even more user-friendly and will enable scientific work as well as applications in various disciplines.

REFERENCES

Breese, J. S. and Heckerman, D. (1996). *Decision-Theoretic Troubleshooting: A Framework for Repair and Experiment*. Morgan Kaufman, San Francisco.

Diez, F. J. and Galan, S. (2003). *Efficient computation for the noisy MAX.* International Journal of Intelligent System, 18, 165–177.

Harmon, P. and King, D. (1985). *Expert Systems.* Wiley, New York.

Hugin (2009). http://www.hugin.com.

Jensen, F. V. (2002). *Bayesian Networks and Decision Graphs.* Springer, New York.

Kern-Isberner, G. (1998). *Characterizing the principle of minimum cross-entropy within a conditional-logical framework.* Artificial Intelligence, New York, 169–208.

Raiffa, H. (1990). *Decision Analysis: Introductory, Lectures on Choice under Uncertainty*. Addison-Wesley, Reading, Massachusetts.

Rödder, W. and Kern-Isberner, G. (2003a). *From Information to Probability - An Axiomatic Approach.* International Journal of Intelligent Systems, 18-4, 383–403.

Rödder, W. and Kern-Isberner, G. (2003b). Self learning or how to make a knowledge base curious about itself. In *German Conference on Artificial Intelligence (KI 2003)*. LNAI 2821, Springer, Berlin, 464–474.

Rödder, W. and Kulmann, F. (2006). *Recall and Reasoning - an Information Theoretical Model of Cognitive Processes.* Information Sciences, 176-17, 2439–2466.

Rödder, W., Reucher, E., and Kulmann, F. (2006). *Features of the Expert-System Shell SPIRIT.* Logic Journal of the IGPL, 14-3, 483–500.

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

Singer, H. (2008). *Maximum Entropy Inference for Mixed Continuous-Discrete Variables.* Faculty of Business Science, University of Hagen, Discussion Paper 432.

Sombe, L. (1990). *Reasoning under Incomplete Information in Artificial Intelligence.* Wiley, New York.

Spirit (2009). http://www.xspirit.de.