Statistica Applicata Vol. 6, n. 3, 1994

# KNOWLEDGE REDUCTION AND INTEGRATION IN PROBABILISTIC EXPERT SYSTEMS

**Zhi Geng[1&4], Chooichiro Asano[1], Minoru Ichimura[2], Hiroshi Kimura[2], Jianhua Guo[3] and Ming Gong[4]**

[1] *Soka University, Japan.* [2] *Okayama University of Science, Japan.* [3] *Qufu Normal University, China.* [4] *Peking University, China.*

*For a rule–based system, we firts discuss how to reduce rules and a knowledge base whithout loss of information on odds and likelihood ratios. Then for a probabilistic expert system, we present a method for integrating inconsistent probabilistic Knowledge from different sources.*

## 1. INTRODUCTION

In this paper, we first discuss how to reduce a knowledge base without loss of information on odds and likelihood ratios. By collapsing and decomposing it, a large contingency table can be reduced into several smaller ones without loss of information on probabilities. Asmussen et al., (1983) and Wermuth et al. (1983) discussed decomposability and collapsibility of multidimensional contingency tables, and then Geng (1989) gave an automatic checking algorithm. However, reasoning methods in many rule–based expert systems are based on odds and likelihood ratios but not on probabilities, such as PROSPECTOR (Duda, et al., 1976). Here we show several results on collapsibility of odds and likelihood ratios. From these results, a knowledge base of odds and likelihood ratios can be reduced without loss of information.

Next we present a method for integrating inconsistent knowledge from different sources. Malvestuto (1986) and Hájek et al., (1992) presented a method for integrating a set of probability distributions which usually is supplied by experts. Their methods are under an assumption that the set must be consistent, that is, distributions in the supplied set must have the same marginal table. However, it is sometimes difficult to check the consistency and we should allow that distributions are inconsistent, especially in the case that knowledge is from different experts and even from different special fields.Here we apply the expectation–maximization (EM) algorithm to integrate the inconsistent distributions supplied by experts.

In section 2, we present definitions and conditions for collapsibilities and show their application to knowledge reduction. Section 3 discusses how to integrate the inconsistent knowledge by the likelihood method.

## 2. COLLAPSIBILITY OF ODDS AND LIKELIHOOD RATIOS

In rule–based expert systems, knowledge is specified as a set of "IF conditions THEN consequence" rules. Conditions in rules can be represented as a conjunction of evidences, say $(E = j) \wedge (F = k)$. Under the condition, the condition includes evidences as few as possible. In this section, we first discuss collapsibilities of the odds and give some results by which rules and evidences in rules can be reduced.

Consider a three dimensional contingency table cross–classified by a consequence H with I categories, and be two evidences E and F with J and K categories respectively. Here we do not limit our discussion in the region of propositional consequence and evidences with logic value. For the $I \times J \times K$ table, let P(H,E,F) denote a probability and P(H|E,F) a conditional probability of H given E and F for H = 1,…,I; E = 1,…, J and F= 1,…, K. Assume that all probabilities are positive.

Define the odds as O(H|E,F) = P(H|E,F) / P(I|E,F) and the marginal odds as O(H|E)=P(H|E)/P(I|E) where P(H|E) is the marginal conditional probability. We say that the odds are collapsibile over F if O(H|E,F) = O (H|E) for all H, E and F.

Suppose that there is a rule "IF $(E = j) \wedge (F = k)$ THEN (H = i) with O(H|E,F)" in a knowledge base. The collapsibililty of odds O(H|E,F) over F means that the evidence F in the rule is not necessary. Thus F can be removed from the conditions.

**Theorem 1.** The necessary and sufficient condition for collapsibility of odds O(H|E,F) over F is the conditional independence of H and F given E, denoted as $H \perp F|E$.

*Proof.* For the sufficiency, we can immediately show O(H|E,F) = O(H|E) from the condition P(H/E,F)=P(H|E). For the necessity, from O(H|E,F) = O(H|E), we have

$$\frac{P(H,F\,|\,E)}{P(I,F\,|\,E)} = \frac{P(H\,|\,E)}{P(I\,|\,E)}.$$

Thus, we get P(F|H,E)=P(F|I,E), which is equivalent to $H \perp F|E$.

In the multi–category case, the collapsibility may be broken by partially grouping the evidence F. Moreover, we define partially marginal odds as $O(H|E, \omega)=P(H|EF \in \omega)|P(I|E, F \in \omega)$ where $\omega$ is a categorical subset of the evidence F.

The collapsible odds O (H|E, F) may be changed after the categories of F are pooled partially, that is, $O(H|E,\omega) \neq O(H|E,F)$ for same $\omega$ even if O(H|E) = O(H|E,F). We say that the odds O(H|E,F) are strongly collapsible over F if $O(H/E,\omega)=O(H|E)$ for all possible subset $\omega \subseteq \{1, ..., K\}$. The strong collapsibility means that the odds remain unchanged no matter how the evidence F is grouped partially.

**Theorem 2**. The necessary and sufficient condition for strong collapsibility of odds O(H|E,F) over F also is $H \perp F|E$.

*Proof.* The necessity can be obtained from theorem 1.For the sufficiency, we have

$$O(H\,|\,E,\omega)=\frac{P(H,\omega\,|\,E)}{P(I,\omega\,|\,E)}=\frac{\Sigma_{F\in\omega}\ P(H,F\,|\,E)}{\Sigma_{F\in\omega}\ P(I,F\,|\,E)}.$$

Since $H \perp F|E$, we can rewrite the above formula as

$$\frac{P(H\,|\,E)\Sigma_{F\in\omega}P(F\,|\,E)}{P(I\,|\,E)\Sigma_{F\in\omega}P(F\,|\,E)}=O(H\,|\,E).$$

Define the likelihood ratios as L(F|H,E) = P(F|H,E)/P(F/I,E) and the marginal likelihood ratios as L(F/H) = P(F/H)/P(F|I) for all E and F and for H = 1,…, I – 1. The odds can be factorized into a likelihood ratio and a prior odds. Now we consider collapsibility of the likelihood ratios.

If L(F|H,E) = L(F|H) for all E, F and H, then we say that the likelihood ratios L(F|H,E) are collapsible over E. The collapsibility is called the modularity property for belief updates by Heckerman (1988). Moreover, we say that L(F|H,E) are strongly collapsible over E if the likelihood ratios remain unchanged no matter how categories of the evidence E are grouped partially, that is, $L(F|H,\omega) = L(F|H)$ for all F and H and for all $\omega \subseteq \{1,...,J\}$ where $L(F|H,\omega)=P(F|H,E\in\omega)/P(F|I,E\in\omega)$. We can reduce the store of likelihood ratios when likelihood ratios are collapsible. Wermuth (1987) and Geng (1992) discussed the collapsibility in terms of relative risks. For the strong collapsibility of likelihood ratios, we can obtain the following theorem from the result on collapsibility of relative risks in Geng (1992).

**Theorem 3.** The necessary and sufficient condition for strong collapsibility of likelihood ratios L(F|H,E) over E is (a)$F \perp E|H$ or (b)$H \perp E|F$ and $H \perp E$, where $H \perp E$ denotes the marginal indipendence of H and E.

From the definitions, the relationship between odds and likelihood ratios can be expressed as

$$O(H|E,F) = L(F|H,E)O(H|E).$$

Odds O(H|E,F) is collapsible over F if and only if L(F|H,E) = 1. Notice that the likelihood ratios L(F|H,E) and the odds O(H|E) are sufficient to determine the odds O(H|E,F). If the likelihood ratios L(F|H,E) are collapsible over E, then we only need L(F|H) and O(H|E). Thus L(F|H,E) stored in a knowledge base can be reduced into L(F|H).

When the likelihood ratios L(F|H,E) is collapsible over E, we have

$$O(H|E,F) = L(F|H)O(H|E) = L(F|H)L(E|H)O(H).$$

Similarly, when the L(E/H,F) is collapsible over F, we also get O(H|E,F) = L(F|H)L(E|H)O(H). In these two cases, the rule "IF (E = i) ^ (F = j) THEN H" can be decomposed into two rules "IF (E = i) THEN H" and "IF (F = j) THEN H", and then evidences E and F can be combined by producting their likelihood ratios L(E|H), L(F|H) and the prior odds O(H).

From theorems 2 and 3, we can see that when $E \perp F|H$, $H \perp E|F$, or $H \perp F|E$ the reason on consequence H from the observed evidences E and/or F just needs the prior odds O(H) and the likelihood ratios L(E|H) and L(F|H).

In many cases, the evidences are considered as propositional ones with only logic values. When the evidence F has only two values (i.e. K = 2), we can obtain the following simpler results.

**Lemma 1.** If K = 2 and $F \perp E|H$, then the following statements are equivalent

(1) $H \perp E$ and $H \not\perp E|F$ and
(2) $(F,E) \perp H$.

*Proof.* It is obvious that (2) implies (1). Here we just show that (1) implies (2). From (1), we have

$$P(H,E,+) = P(H,+,+)P(+,E,+)$$

and

$$P(H,E,F) = \frac{P(H,+,F)P(+,E,F)}{P(+,+,F)}.$$

Since

$$P(H,E,+) - P(H,E,1) = \frac{(P(H,+,+) - P(H,+,1))(P(+,E,+) - P(+,E,1))}{1 - P(+,+,1)},$$

we get

$$P(H,E,1)+P(H,E,+)P(+,+,1) = P(H,+,+)P(+,E,1)+P(H,+,1)P(+,E,+).$$

So

$$P(H,E,1)-P(H,+,1)P(+,E,+) = P(H,+,+)(P(+,E,1)-P(+,+,1)P(+,E,+)).$$

Since P(H,E,1) = P(H,+,1)P(+,E,1)/P(+,+,1), we have

$$\frac{P(H,+,1)(P(+,E,1)-P(+,+,1)P(+,E,+))}{P(+,+,1)} = P(H,+,+)(P(+,E,1)-P(+,+,1)P(+,E,+)).$$

Multiplying P(+,+,1) and moving the left term to right, we get

$$(P(+,E,1)-P(+,+,1)P(+,E,+))(P(H,+,+)P(+,+,1)-P(H,+,1)) = 0.$$

That is,

$$P(+,E,1) = P(+,+,1)P(+,E,+), \quad \forall E$$

or

$$P(H,+,1) = P(+,+,1)P(H,+,+), \quad \forall H.$$

In the case of K = 2, the above two equations are equivalent to $F \perp E$ and $F \perp H$ respectively.

If $F \perp E$, then

$$P(H,E,F) = \frac{P(H,+,F)P(+,E,F)}{P(+,+,F)} = P(H,+,F)P(+,E,+).$$

The above equation is equivalent to $(F,H) \perp E$, which implies $F \perp E|H$. From the condition of the lemma, we obtain $F \not\perp E$. Thus we must have $F \perp H$ and so

$$P(H,E,F) = \frac{P(H,+,F)P(+,E,F)}{P(+,+,F)} = P(H,+,+)P(+,E,F).$$

We have shown (2).

**Theorem 4.** When K = 2, the necessary and sufficient condition for strong collapsibility of likelihood ratios L(F|H,E) over E is (a) $F \perp E|H$ or (b) $(F,E) \perp H$.

*Proof.* The sufficiency can be immediately obtained from theorem 3 since $(F,E) \perp H$ implies both $H \perp E|F$ and $H \perp E$. The necessity can be shown from theorem 3 and lemma 1.

Notice that the teorem 4 does not hold for K > 2.

On the order hand of knowledge reduction, we can also consider strong collapsibility as a criterion for recategorizing an evidence. A categorical subset of an evidence can be pooled into a new single category if the likelihood ratio or the odds is strongly collapsible conditionally on the subset, that is, it remains unchanged no matter how categories in the subset are pooled partially. In this way, we can obtain a recategorized evidence with more concise categories. For each category of recategorized evidence, the conditional likelihood ratio or odds remains unchanged by further sub–categorizing the category.

Moreover we say that the likelihood ratios O(H|E,F) are weakly collapsible over, F if O(H|E,F) ) ≥ 1 (or ≤ 1) for all F implies O(H|E) ≥1 (or ≤ 1) for all H and E.

**Theorem 5.** Let $LOW(H,E) = \min_F O(H|E,F)$ and $UP(H,E) = \max_F O(H|E,F)$. Then $LOW(H|E) \le O(H/E) \le UP(H,E)$.

*Proof.* From the definition, we have

$$H|E = \frac{P(H|E)}{P(I|E)} = \frac{\sum_F O(H|E,F)P(I,E,F)}{\sum_F P(I,E,F)}$$

$$\ge \frac{LOW(H,E)\sum_F P(I,E,F)}{\sum_F P(I,E,F)} = LOW(H,E).$$

Similary, we can obtain $O(H|E) \le UP(H,E)$.

Thus we see that the likelihood ratios always are weakly collapsible. Unlike odds ratios, it is impossible that the likelihood ratios have a dramatic change when some evidence is ignored.

## 3. INTEGRATION OF INCONSISTENT KNOWLEDGE

In this section, we consider how to integrate the inconsistent knowledge from different sources. Inconsistency may occur in knowledge from one expert or among different experts. Here we expect that experts supply frequencies as knowledges but not probabilities since frequencies contain not only information on probabilities but also information on weights. Frequencies can represent degrees of belief on different experts to different marginal tables. If an expert has supplied only probabilities, we need to weigh them by multiplying a factor which reflects the expert's degree.

Without loss of generality, we only describe the case with two knowledge sources, say two experts supply their knowledges. Consider a knowledge base denoted by a multidimensional contingency table with a variable set $\Omega$. Suppose the first expert supply a class of knowledge configurations, $T_1 = \{t_1^1, \ldots, t_n^1\}$ where $t_j^1$ is a subset of variables, i.e. $t_j^1 \subseteq \Omega$ for all j. And for each configuration $t \in T_1$, he supplies frequencies, $N_t^1(i_t)$ for all $i_t$ where $i_t$ is an index of the table cross–classified by variables in t. Notice that a knowledge configuration may be contained by another, i.e. $t_j^1 \subseteq t_k^1$ and that total frequencies may be different, i.e. $N_t^1(+) \neq N_s^1(+)$ for some t and $s \in T_1$ where $N_t^1(+)$ denotes the sum of $N_t^1(i_t)$ over $i_t$. Similarly, the second expert supplies a knowledge class $T_2$ and frequencies $N_t^2(i_t)$ for all $i_t$ and $t \in T_2$. We say that knowledges are inconsistent in an expert if the frequencies from the same expert are inconsistent in a marginal table, that is, there are two configurations $t_j^r$ and $t_k^r$ in some $T_r$ such that $t_j^r \cap t_k^r = s$ and $N_{t_j^r}^r(i_s) \neq N_{t_k^r}^r(i_s)$ for some $i_s$, where $N_t^r(i_\Phi)$ means the total frequency of $N_t^r(i_t)$. Knowledges are inconsistent between expert if the frequencies from different experts are inconsistent, that this, tehere are two configurations $t_k^1$ and $t_k^2$ such that $t_j^1 \cap t_k^2 = s$ and $N_{t_j^1}^1(i_s) \neq N_{t_k^2}^2(i_s)$ for some $i_s$.

Let T = $T_1$ U $T_2$ which may contain the duplicates. We see frequencies from two experts, $N_t(i_t)$ for all $i_t$ and $t \in T$, as observations from the same population. Assume that the information on frequencies in any configuration does not depend on variables outside the configuration, that is, for any t in T, $N_t(i_t)$ indeed only presents the marginal information on variables in t. Then the log likelihood function is defined as

$$L \propto \sum_{t \in T} \sum_{i_t} N_t(i_t) \log P(i_t)$$

wehere $P(i_t)$ is a probability in the cell $i_t$. This is the same as the log likelihood function of incomplete data under an assumption that missing values are missing at random. Therefore we can find the maximum likekihood estimates $\hat{P}(i_\Omega)$ using the EM algorithm, see Little et al. (1987) and Fuchs (1982). Thus the estimates $\hat{P}(i_\Omega)$ can be used as the integrated knowledge.

For some special cases, we can decompose an estimation problem of a large table into that of several smaller tables. For further details, see Geng (1988) and Geng et al. (1994). By decomposability, we can reduce the above EM calculation to several local calculation.

## ***ACKNOWLEDGEMENTS***

*The first author would like to thank Professor A. Onishi, the vice–president of Soka University, for his kind support for this research during the period of the author's visiting Soka University, Japan. This research was partially supprted by the National Natural Science Foundation of China.*

## REFERENCES

Asmusses, S. and Edwards, D. G., 1983, Collapsibility and response variables in contingency tables, *Biometrika,* 70, 567–578.

Duda, R., Hart, P., Nilsson, N. and Sutherland, G., 1976, Subjective Bayesian methods for rule–based inference systems, *Proceedings of the National Computer Conference,* 1075–1082.

Fuchs, C., 1982, Maximum likelihood estimation and model selection in contingency tables with missing data, *J. Am. Statist. Assoc.,* 77,270–278.

Geng, Z., 1988, Multidimensional contingency tables with missing data, *Commun. Statist.–Theory and Meth.,* 17, 4137–4146.

Geng, Z., 1989, Decomposability and collapsibility for log–linear models, *Applied Statisics,* 38, 189–197.

Geng, Z., 1992, Collapsibility of relative risk in contingency tables with a response variable, *J. Royal Statist. Soc.,* B54, 585–593.

Geng, Z., Asano, Ch., Ichimura, M., and Kimura, H., 1994, Decomposability and collapsibility for contingency tables with missing data. *To appear in Applied Statistics,* Vol. 43, No. 3.

Hâjek, P., Havrânek, T. and Jirousek, R., 1992, *Uncertain Information Processing in Expert Systems,* CRC Press, Boca Raton.

Heckerman, D.E., 1988, An axiomatic framework for belief updates, *Uncertainty in Artificial Intelligence 2,* Lemmer, J.F. and Kanal, L. N. eds, North Holland, 11–22.

Little, R. J. A. and Rubin, D.B. 1987 *Statistical Analysis with Missing Data,* John Wiley, New York.

Malvestuto, F. M., 1986, Statistical treatment of the information content of a database, *Inform. Syst.,* 11, 211–223.

Wermuth, N., 1987, Parametric collapsibility and the lack of moderating effects in contingency tables with a dichotomous response variable, *J. Royal Statist. Soc.* B, 49 353–364.

Wermuth, N. and Lauritzen, S. L., 1983, Graphical and recursive models for contingency tables, *Biometrika,* 70, 537–552.

## *RIASSUNTO*

*Innanzitutto discutiamo su come ridurre le regole e una base della conoscenza senza perdita di informazione sui rapporti probabilistici e di verosimiglianza in un sistema basato su regole. Quindi, presentiamo un metodo per integrare la conoscenza probabilistica inconsistente da fonti diverse in un sistema esperto probabilistico.*