INCONSISTENT KNOWLEDGE INTEGRATION IN A PROBABILISTIC MODEL†

Radim Jiroušek[1,2], Jiří Vomlel[1]

[1]Department of Decision-Making Theory
Institute of Information Theory and Automation
Pod vodárenskou věží 4,
182 08 Praha 8, Czech Republic

[2]Department of Information and Knowledge Engineering
Faculty of Informatics and Statistics
University of Economics
Náměstí W. Churchilla 4,
130 67 Praha 3, Czech Republic

INTRODUCTION

In probabilistic models of knowledge based systems, the knowledge base is usually represented by a multidimensional probability distribution (Hájek et. al., 1992). Analogously, oligodimensional distributions (i.e. distributions of dimensionality cca 1 - 5) can be considered partial knowledge. Within this framework the knowledge integration process corresponds to constructing a multidimensional distribution with the given marginals. For this purpose the well known Iterative Proportional Fitting (IPF) procedure can be used.

It is known that IPF converges if the set of oligodimensional distributions is consistent, i.e. if there exists at least one distribution with the given marginals. In opposite case the procedure tends to cycle. Nevertheless, in practical situations one can wish to get some representative of the input knowledge even in cases when it is inconsistent (especially because of the fact that no procedure for consistency testing is known).

Therefore, we are looking for a modification of IPF that meets the following **basic requirements**:

†The research was financially supported by the grants: GA AVČR 27564, GA ČR 201/93/0781, 201/94/0471, EC CIPA3511CT930053.

(i) if the input oligodimensional distributions are consistent, the procedure converges to the same distribution as the original IPF,

(ii) the procedure converges even in inconsistent situations,

(iii) the limit distribution is independent of the ordering of the input oligodimensional distributions,

(iv) and the marginals of the resulting distribution "are close" to the input oligodimensional distributions.

In this paper, two approaches are presented that seem to meet these requirements. At present we are testing their behaviour and this paper summarizes some of achieved experimental results.

PROBABILISTIC MODEL

Let V denote a finite index set of a system $\{X_i\}_{i\in V}$ of finite-valued variables for which we are looking for a joint probability distribution $Q((X_i)_{i\in V})$. Further, let C_k for $k=1,\ldots,K$ be subsets of this index set V. For each C_k we assume that a probability distribution P_k defined for variables $(X_i)_{i\in C_k}$ is given. This system of oligodimensional distributions will be denoted by Θ:

$$\Theta = \{P_1,\ldots,P_K\}.$$

Considering any joint probability distribution $Q((X_i)_{i\in V})$ and a subset $C\subseteq V$, the symbol Q^C denotes the marginal distribution of Q defined for variables $(X_i)_{i\in C}$. For any system of oligodimensional distributions $\Theta=\{P_1,\ldots,P_K\}$ defined for variables $\{X_i\}_{i\in C_1},\ldots,\{X_i\}_{i\in C_K}$ respectively we can specify the system

$$\Pi(\Theta)=\{R(X_i)_{i\in V} : R^{C_k}=P_k, \forall_k=1,\ldots,K\}$$

of all the joint probability distributions consistent with all the oligodimensional distributions from Θ. Generally, it is a hard problem to determine properties of the system $\Pi(\Theta)$ when only distributions from Θ are known. It is also a hard task to construct some distribution from $\Pi(\Theta)$. A partial answer to these problems is given by the following assertion proven by Csiszár (1975).

Assertion. *Let Q_0 be the joint uniform probability distribution defined for all the variables $(X_i)_{i\in V}$. The sequence of joint distributions $Q_1,Q_2,\ldots$ defined by the formula*

$$Q_n = Q_{n-1}\cdot\frac{P_j}{Q_{n-1}^{C_j}} \qquad \text{for} \quad j=((n-1) \bmod K)+1 \tag{1}$$

is convergent iff $\Theta=\{P_1,\ldots,P_k\}$ is strongly consistent, i.e. $\Pi(\Theta)\neq\emptyset$. In the case of convergence

$$Q=\lim_{n\to\infty} Q_n \in \Pi(\Theta),$$

and

$$H(Q)=\max_{R\in\Pi(\Theta)} H(R),$$

where $H(R)$ denotes the Shannon entropy of the distribution R

$$H(R)=-\sum_x R(x)\log R(x).$$

Remark. The computational process defined by the formula (1) is known as the *Iterative Proportional Fitting* procedure (IPF). As it is explained in (Cziszár, 1975), the distribution Q_n is an *I-projection* of Q_{n-1} into the class of distributions $\Pi(\{P_j\}) = \{R(X_i)_{i\in V} : R^{C_j} = P_j\}$ in the following sense

$$Q_n \in \Pi(\{P_j\}) \quad \& \quad I(Q_n \parallel Q_{n-1}) = \min_{R\in\Pi(\{P_j\})} I(R \parallel Q_{n-1}),$$

where $I(P \parallel R)$ denotes the Kullback-Leibler divergence (I-divergence) defined

$$I(P \parallel R) = \sum_x P(x) \log \frac{P(x)}{R(x)}. \tag{2}$$

INCONSISTENT SITUATION

What can be said about the IPF process when the system Θ of oligodimensional distributions is not consistent, i.e. when $(\Pi(\Theta) = \emptyset)$?

In the sequel we shall sometimes distinguish two situations. We shall say that a system $\Theta = \{P_1, \ldots, P_k\}$ is *weakly consistent* if for any couple of different indices $k, \ell \in \{1, \ldots, K\}$

$$P_k((X_i)_{i\in(C_k\cap C_\ell)}) = P_\ell((X_i)_{i\in(C_k\cap C_\ell)}),^{\ddagger}$$

or, equivalently

$$P_k^{C_k\cap C_\ell} = P_\ell^{C_k\cap C_\ell}.$$

It is evident that the weak consistency is a necessary condition for a system Θ to be consistent (further we shall also say *strongly consistent*). Nevertheless, as it will be seen from examples, it is not a sufficient condition. All the examples presented in this paper will be examples of inconsistent (i.e. strongly inconsistent) systems of oligodimensional distributions that will be weakly consistent.

After a number of computational experiments with the classical IPF procedure applied to both weakly consistent and inconsistent systems of oligodimensional distributions we believe that only two possibilities can happen.

Conjecture 1. *Let Q_0 be the joint uniform probability distribution defined for all the variables $(X_i)_{i\in V}$. For any system of strongly inconsistent oligodimensional distributions application of the IPF procedure (formula (1)) leads to one of the following two situations:*

- *The computation fails because there exits $1 < j \leq K$ such that the distribution P_j is not dominated by $Q_{j-1}^{C_j}$ (i.e. there exists an event x such that $P_j(x) > 0$ and $Q_{j-1}^{C_j}(x) = 0$), and thus the distribution Q_j is not defined.*
- *The sequence $Q_1, Q_2, \ldots$ tends to cycle. More precisely; for all $1 \leq j \leq K$ the subsequences $Q_j, Q_{j+K}, Q_{j+2K}, \ldots, Q_{j+nK}, \ldots$ are convergent. All the limit distributions for all these subsequences differ each from other.*

Let us illustrate these two possibilities on simple examples.

Example 1. Consider three two-dimensional distributions $\{P_1(X_1, X_2),\ P_2(X_2, X_3),\ P_3(X_1, X_3)\}$ given in Table 1.

---

‡For $C_k \cap C_\ell = \emptyset$ the condition is always fulfilled.

Table 1. Input distributions $\{P_1,\ P_2,\ P_3\}$

P_1	$X_2=1$	$X_2=2$	P_2	$X_3=1$	$X_3=2$	P_3	$X_3=1$	$X_3=2$
$X_1=1$	$0.5-\varepsilon$	$0+\varepsilon$	$X_2=1$	$0.5-\varepsilon$	$0+\varepsilon$	$X_1=1$	$0+\varepsilon$	$0.5-\varepsilon$
$X_1=2$	$0+\varepsilon$	$0.5-\varepsilon$	$X_2=2$	$0+\varepsilon$	$0.5-\varepsilon$	$X_1=2$	$0.5-\varepsilon$	$0+\varepsilon$

Taking $\varepsilon=0$, the computation fails because the distribution P_3 is not dominated by $Q_2^{C_3}$. Namely, for $x_{12}=(X_1=1, X_3=2)$ $P_3(x_{12})>0$ and $Q_2^{C_3}(x_{12})=0$ as well as for $x_{21}=(X_1=2, X_3=1)$ $P_3(x_{21})>0$ and $Q_2^{C_3}(x_{21})=0$, and thus the distribution Q_3 is not defined. The respective distributions Q_1 and Q_2 are in Table 2.

The other situation occurs for $\varepsilon=0.1$. In this case the sequence of distribution $Q_1, Q_2, \ldots$ "converges" to the cycle of the three distributions shown in Table 3.

Table 2. Application of IPF to the distributions from Table 1 with $\varepsilon=0$.

X_1	X_2	X_3	Q_0	Q_1	Q_2	Q_3
1	1	1	0.1250	0.2500	0.5000	0.0000
1	1	2	0.1250	0.2500	0.0000	-
1	2	1	0.1250	0.0000	0.0000	0.0000
1	2	2	0.1250	0.0000	0.0000	-
2	1	1	0.1250	0.0000	0.0000	0.0000
2	1	2	0.1250	0.0000	0.0000	-
2	2	1	0.1250	0.2500	0.0000	0.0000
2	2	2	0.1250	0.2500	0.0000	-

Table 3. Application of IPF to the distributions from Table 1 with $\varepsilon=0.1$. Their ordering was $\{P_1, P_2, P_3\}$

X_1	X_2	X_3	Q_0	Q_1	Q_2	Q_3	...	Q_{19+3n}	Q_{20+3n}	Q_{21+3n}
1	1	1	0.1250	0.2000	0.3200	0.0941		0.1562	0.2439	0.1000
1	1	2	0.1250	0.2000	0.0800	0.2000		0.2438	0.1000	0.1562
1	2	1	0.1250	0.0500	0.0200	0.0059		0.0000	0.0000	0.0000
1	2	2	0.1250	0.0500	0.0800	0.2000		0.1000	0.1562	0.2438
2	1	1	0.1250	0.0500	0.0800	0.2000		0.1000	0.1562	0.2438
2	1	2	0.1250	0.0500	0.0200	0.0059		0.0000	0.0000	0.0000
2	2	1	0.1250	0.2000	0.0800	0.2000		0.2438	0.1000	0.1562
2	2	2	0.1250	0.2000	0.3200	0.0941		0.1562	0.2439	0.1000

Remark. Let us stress that even for weakly consistent systems of oligodimensional distributions the systems of limit distributions may differ when the distributions in Θ are reordered.

CONSERVATIVE MODIFICATION OF IPF

The following modification of IPF procedure computes at each step of its process a convex mixture of the usually computed I-projection and of the distribution corresponding to the previous step

$$Q_n = (1-\alpha_n)\cdot Q_{n-1} + \alpha_n \cdot Q_{n-1} \cdot \frac{P_j}{Q_{n-1}^{C_j}} \qquad \text{for } j = ((n-1) \bmod K)+1, \tag{3}$$

for a monotonous sequence $\{\alpha_n\}$ with $\lim_{n\to\infty} \alpha_n = 0$.

This modification has been tested with several types of input sets of oligodimensional distributions and with several different types of sequences of α_n as, for example,

$$\alpha_n = \frac{1}{1+n}, \qquad \alpha_n = \frac{1}{\sqrt{1+n}}, \qquad \alpha_n = \frac{1}{\pi}(\frac{\pi}{2} - arctg\ n).$$

The results achieved lead us to the following conjecture.

Conjecture 2. *If $\sum_{n=1}^{\infty} \alpha_n = +\infty$ then the conservative modification of IPF (formula (3)) meets the properties (i), (ii) and (iii) of the basic requirements. The limit distribution Q minimizes the sum of the total variances:*

$$\sum_{k=1}^{K} V(Q^{C_k}, P_k) = \min_R \sum_{k=1}^{K} V(R^{C_k}, P_k),$$

where

$$V(R,P) = \sum_x |R(x) - P(x)|. \tag{4}$$

Remark. In contrast to the I-divergence, minimization of the sum of the total variances does not lead to a unique distribution but usually a whole convex set of distributions minimize the value $\sum_{k=1}^{K} |R^{C_k}(x) - P_k(x)|$ (see also Table 8).

Example 2. Let us now consider another weakly consistent system of two-dimensional distributions $\{P_1(X_1,X_2),\ P_2(X_2,X_3),\ P_3(X_1,X_3)\}$ given in Table 4. Results of application of the conservative modification of IPF to this system are shown in Tables 5 and 6. They contain results achieved for orderings P_1, P_2, P_3 and P_3, P_2, P_1 respectively. In both cases the sequence

$$\alpha_n = \frac{1}{1+i}, \qquad \text{for} \quad i = ((n-1) \text{ div } K) + 1$$

was used.

Table 4. Input distributions $\{P_1,\ P_2,\ P_3\}$

P_1	$X_2=1$	$X_2=2$	P_2	$X_3=1$	$X_3=2$	P_3	$X_3=1$	$X_3=2$
$X_1=1$	0.02	0.38	$X_2=1$	0.13	0.17	$X_1=1$	0.24	0.16
$X_1=2$	0.28	0.32	$X_2=2$	0.67	0.03	$X_1=2$	0.56	0.04

Table 5. Application of the conservative modification of IPF to the distributions from Table 4. Ordering $\{P_1, P_2, P_3\}$.

X_1	X_2	X_3	Q_0	Q_1	Q_2	Q_3	Q_4	...	Q_{10001}
1	1	1	0.1250	0.0675	0.0557	0.0494	0.0358		0.0000
1	1	2	0.1250	0.0675	0.0624	0.0647	0.0469		0.0586
1	2	1	0.1250	0.1575	0.2546	0.2258	0.2411		0.2838
1	2	2	0.1250	0.1575	0.0866	0.0898	0.0959		0.0581
2	1	1	0.1250	0.1325	0.1093	0.1448	0.1584		0.1678
2	1	2	0.1250	0.1325	0.1226	0.0735	0.0804		0.0732
2	2	1	0.1250	0.1425	0.2304	0.3051	0.2958		0.3585
2	2	2	0.1250	0.1425	0.0784	0.0470	0.0456		0.0000

Table 6. Application of the conservative modification of IPF to the distributions from Table 4. Ordering $\{P_3, P_2, P_1\}$.

X_1	X_2	X_3	Q_0	Q_1	Q_2	Q_3	Q_4	...	Q_{10001}
1	1	1	0.1250	0.1225	0.0858	0.0475	0.0450		0.0000
1	1	2	0.1250	0.1025	0.1010	0.0559	0.0599		0.0585
1	2	1	0.1250	0.1225	0.1875	0.2377	0.2251		0.2838
1	2	2	0.1250	0.1025	0.0600	0.0761	0.0815		0.0581
2	1	1	0.1250	0.2025	0.1417	0.1640	0.1759		0.1677
2	1	2	0.1250	0.0725	0.0715	0.0827	0.0641		0.0733
2	2	1	0.1250	0.2025	0.3100	0.2957	0.3172		0.3585
2	2	2	0.1250	0.0725	0.0425	0.0405	0.0314		0.0000

METHOD OF ITERATIVE AVERAGES

This procedure was proposed by F. Matúš (1994). It is based on the idea to prevent cycling by computing an average of all K I-projections at each step. More precisely, this procedure consists in computation of the distributions

$$Q_n = \frac{1}{K} \sum_{j=1}^{K} Q_{n-1} \cdot \frac{P_j}{Q_{n-1}^{C_j}}. \tag{5}$$

Conjecture 3. *The method of iterative averages converges to the same limit distribution as the conservative modification of IPF.*

Example 3. Applying the method of iterative averages to the three two-dimensional distributions $\{P_1(X_1, X_2),\ P_2(X_2, X_3),\ P_3(X_1, X_3)\}$ from Table 4 we get the limit distribution already at 41 steps. As it can be seen from Table 7, the limit distribution really does not differ from the results achieved by the conservative modification of IPF.

Table 7. Application of the method of iterative averages to the distributions from Table 4.

X_1	X_2	X_3	Q_0	Q_1	Q_2	Q_3	...	Q_{41}
1	1	1	0.1250	0.0650	0.0345	0.0200		0.0000
1	1	2	0.1250	0.0583	0.0471	0.0464		0.0588
1	2	1	0.1250	0.2150	0.2507	0.2648		0.2839
1	2	2	0.1250	0.0950	0.0778	0.0704		0.0579
2	1	1	0.1250	0.1617	0.1648	0.1640		0.1680
2	1	2	0.1250	0.0817	0.0718	0.0718		0.0730
2	2	1	0.1250	0.2583	0.3219	0.3462		0.3584
2	2	2	0.1250	0.0650	0.0314	0.0164		0.0000

COMPARISONS AND CONCLUSIONS

What can be said as a conclusion at the current state of research? We have presented two iterative methods converging even for inconsistent systems of oligodimensional distributions. As it was stated in Conjecture 3, we believe that both of them yield the same probability distribution. Another distribution with similar properties can be received by application of the classical IPF procedure to the inconsistent system of distributions, and then, if the process starts to cycle, one can simply compute the average (barycentre) of the respective "limit" distributions. A disadvantage of this process is that it violates the condition (iii) of the basic requirements and that it cannot be sometimes used (see the first situation of Conjecture 1). Nevertheless, the results achieved in this way do not differ too much from those received by the conservative modification of IPF. An example supporting this statement is presented in Table 8 containing distributions computed for the system of distributions given in Table 4.

In Table 8, the following symbols are used:

$Q_{(B_1)}$ – the barycentre distribution of the limit cycle, when clasical IPF procedure is applied to the ordering $\{P_1, P_2, P_3\}$,

$Q_{(B_2)}$ – the barycentre distribution of the limit cycle, when clasical IPF procedure is applied to the ordering $\{P_3, P_2, P_1\}$,

$Q_{(C)}$ – the distribution computed with the conservative modification of IPF,

$Q_{(A)}$ – the distribution computed with the method of iterative averages,

$Q_{(I_{min})}$ – the distribution minimizing the sum of Kullback-Leibler divergences from the input oligodimensional distributions.

Table 8. Comparison of the results achieved with the input system given in Table 4.

X_1	X_2	X_3	$Q_{(B_1)}$	$Q_{(B_2)}$	$Q_{(C)}$	$Q_{(A)}$	$Q_{(I_{min})}$
1	1	1	0.0000	0.0000	0.0000	0.0000	0.0000
1	1	2	0.0528	0.0578	0.0586	0.0588	0.0477
1	2	1	0.2874	0.2836	0.2838	0.2839	0.2898
1	2	2	0.0588	0.0577	0.0581	0.0579	0.0552
2	1	1	0.1682	0.1709	0.1678	0.1680	0.1695
2	1	2	0.0749	0.0703	0.0732	0.0730	0.0713
2	2	1	0.3579	0.3597	0.3585	0.3584	0.3665
2	2	2	0.0000	0.0000	0.0000	0.0000	0.0000
$\sum_{k=1}^{K} V(Q^{C_k}, P_k)$			0.440	0.440	0.440	0.440	0.440
$\sum_{k=1}^{K} I(Q^{C_k} \parallel P_k)$			0.0726	0.0736	0.0744	0.0744	0.0711

REFERENCES

Csiszár, I., 1975, I-divergence geometry of probability distribution and minimization problems, *Ann.Prob.* 3: 146-158.

Hájek, P., Havránek, T., and Jiroušek, R., 1992, "Uncertain Information Processing in Expert Systems," CRC Press, Inc., Boca Raton.

Matúš, F., 1994, Personal communication.