TRANSFER OF KNOWLEDGE IN CHEMICAL EQUIPMENT RELIABILITY

František BABINEC and Mirko DOHNAL

Department of Chemical Machines, Faculty of Mechanical Engineering, Technical University, 619 69 Brno

Received November 1, 1988
Accepted January 24, 1989

The problem of transformation of data on the reliability of chemical equipment obtained in particular conditions to other equipment in other conditions is treated. A fuzzy clustering algorithm is defined for this problem. The method is illustrated on a case study.

Reliability is among today's most important problems that have to be solved for intensification of any existing process and, in particular, for any intended new investment action. In fact, the public has been alerted by the past series of serious accidents, with the results that conditions under which new productions are permitted are being made more and more stringent[1].

Securing performance reliability and safety is a complex problem. An exceedingly important component in its solving is knowledge handling. It is clear now that it does not suffice to employ information that can be quantified numerically by conventional methods[2]; a wealth of other information, such as cannot be simply transformed into numbers, is then lost.

Additional loss of information occurs as the numerical information is being processed for storage within the framework of some particular information structure, such as a data base or part of an expert base in an expert system. Since information concerning reliability and safety is drastically insufficient on its own, it is desirable that loss of such information be kept as low as possible. Fortunately, methods developed for unconventional knowledge handling are also applicable to reliability[3].

Experience gained in the handling of data related to reliability (not only in chemistry but also, e.g., for ball bearings, packings, presses, etc.) shows that a crucial role in reliability is played by the field of problems which in artificial intelligence have been solved by analogy[4].

In practice, this implies that reliability data obtained for some equipment in particular conditions have to be transformed to other equipment and other conditions. Naturally, the less extensive the transformation, the more reliable the result. For example, if data on the service life of a heat exchanger under precisely specified conditions are employed for making an estimate under insignificantly altered conditions, it is reasonable to assume that the service life prediction will be sufficiently likely. On the other hand, if the transfer involves significant changes both in the equipment and in the conditions, the prediction likelihood will be appreciably lower.

This particular problem is the subject of the present work. A fuzzy clustering algorithm has to be defined for its treatment.

Fuzzy Clustering Algorithm

Clustering has actually become a classical approach to data processing[5]. Since vague information is often to be processed, it is convenient to use fuzzy clustering algorithms, a number of which exist[6].

In the following text, an algorithm based on the testing of the internal consistency of a fuzzy expert base will be used.

An expert base is a set of conditional statements

$$\text{if } \mathbf{A}_i \text{ then } \mathbf{B}_i \quad i = 1, 2, \ldots, n \,, \tag{1}$$

where $\mathbf{A}_i$ is an m-dimensional fuzzy set of the type

$$\mathbf{A}_1 \text{ and } \mathbf{A}_2 \text{ and} \ldots \text{and } \mathbf{A}_m \tag{2}$$

(for details see, e.g., ref.[7]).

Now assume that n records on failures exist in the expert base of the form (*1*). Delete the r-th record, i.e. the record in the form

$$\text{if } \mathbf{A}_r \text{ then } \mathbf{B}_r \tag{3}$$

from the expert base. We inquire the reduced data base, putting the question which is identical with the extracted failure record. The expert base

$$\begin{gathered} \text{if } \mathbf{A}_1 \text{ then } \mathbf{B}_1 \\ \vdots \\ \text{if } \mathbf{A}_{r-1} \text{ then } \mathbf{B}_{r-1} \\ \text{if } \mathbf{A}_{r+1} \text{ then } \mathbf{B}_{r+1} \\ \vdots \\ \text{if } \mathbf{A}_n \text{ then } \mathbf{B}_n \end{gathered} \tag{4}$$

will answer the question represented by the fuzzy set $\mathbf{A}_r$ (see relation (*3*)). The reply to such a question should somehow be similar to the fuzzy set $\mathbf{B}_r$ of the above record.

If this is not the case, then either there is an error in the expert base, or the failure is atypical.

A problem of internal consistency testing lies in the evaluation of similarity. Actually, fuzzy similarity does not enable us to choose one of two possibilities of the form

$$\boldsymbol{B}_r = \boldsymbol{B}_r \text{ reply}$$
$$\boldsymbol{B}_r \neq \boldsymbol{B}_r \text{ reply} . \tag{5}$$

As usual in fuzzy set theory, we must admit that membership is partial in the two alternatives rather than unique.

The similarity definition given below can be employed for applications in reliability analysis and accident analysis.

The first type of similarity is defined as follows:

$$S1 = NR1(\boldsymbol{B}_r \cap \boldsymbol{B}_{r,reply}) , \tag{6}$$

where NR1 is the first numerical representation. The second type of similarity is defined as follows:

$$S2 = |NR2(\boldsymbol{B}_r) - NR2(\boldsymbol{B}_{r,reply})| , \tag{7}$$

where NR2 is the second numerical representation of the fuzzy set.

NR2($\boldsymbol{A}$) represents the x-coordinate of the centre of gravity of the area bounded by the graph of the degree of membership $m_A(x)$ of the fuzzy set $\boldsymbol{A}$. NR1($\boldsymbol{A}$) is the value of the variable x for which the degree of membership $m_A(x)$ has the maximum value

$$NR1(\boldsymbol{A}) = \max_x [m_A(x)] . \tag{8}$$

For details see ref.[8].

Data Transfer

Assume for simplicity that accident records are described by means of two variables only, X_1 and X_2. This makes possible a graphic illustration of the solution process. Let each one-dimensional fuzzy set have the following graph of membership degree.

The expert base can be divided into four clustering regions. Each clustering region $\boldsymbol{C}^{ij}$ is characterized by the product of two fuzzy sets $\boldsymbol{C}^i$ and $\boldsymbol{C}^j$ (see Figs 1 and 2):

$$\boldsymbol{C}^{11} = \boldsymbol{C}_1^1 \times \boldsymbol{C}_2^1$$
$$\boldsymbol{C}^{12} = \boldsymbol{C}_1^1 \times \boldsymbol{C}_2^2$$
$$\boldsymbol{C}^{21} = \boldsymbol{C}_1^2 \times \boldsymbol{C}_2^1 \tag{9}$$
$$\boldsymbol{C}^{22} = \boldsymbol{C}_1^2 \times \boldsymbol{C}_2^2$$

In Fig. 2, the region $\mathbf{C}^{11}$ is hatched as an example of those pairs of values of the variables X_1 and X_2 whose degree of membership in the clustering region $\mathbf{C}^{11}$ is one.

The four fuzzy sets $\mathbf{C}^{11}$, $\mathbf{C}^{12}$, $\mathbf{C}^{21}$ and $\mathbf{C}^{22}$ are shown as polygon functions on the corresponding axes (Fig. 1). It is apparent that on the X_1 axis, two sets overlap, having an intersection. On the X_2 axis, the two sets appear as disjoint. Given by the particular problem, their mutual position is of no consequence in the following treatment.

Since the accident records (see relation (3)) are assumed to be two-dimensional only, the statements by the expert base can be represented in Fig. 2 by rectangles showing the regions of nonzero degrees of membership. One such record is shown in Fig. 2 by means of the fuzzy sets $\mathbf{N}_1$ and $\mathbf{N}_2$. It is also clear that the fuzzy set $\mathbf{N}$ falls in the clustering region $\mathbf{C}^{12}$.

From the expert base standpoint, the clustering regions $\mathbf{C}^{ij}$ can be regarded as questions put to the base. The clustering regions are to be chosen with regard to the particular knowledge transfer. Hence, the shape of the clustering regions is determined by the source and target of the knowledge transfer.

Accidents that belong to a clustering region are members of this region. It can also happen that one and the same accident will belong to several clustering regions with membership degrees lower than one. In our case this is due to the fact that the fuzzy sets $\mathbf{C}_1^1$ and $\mathbf{C}_1^2$ overlap (Fig. 2).

For illustration of the data transfer, assume that all the four clustering regions have been put as questions and no accident has been activated by the region $\mathbf{C}^{22}$.

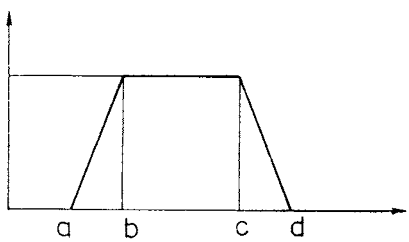

FIG. 1

Approximation of the membership function by a polygon linear dependence

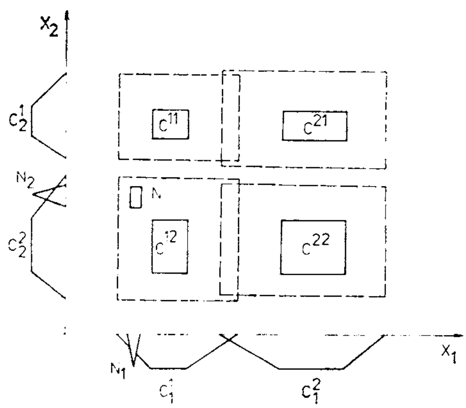

FIG. 2

Graphic interpretation of knowledge transfer

Practically, the latter implies that no accident has as yet occurred under the conditions considered; also this can imply that the equipment has not yet been operated under these conditions. At any rate this means that primary data are lacking for this clustering region $\mathbf{C}^{22}$.

The so-called delta cluster can be conveniently employed for describing the data transfer. This delta cluster is based on so-called delta statements, which can be written as follows:

$$\text{if } \Delta x_{ij} \text{ then } \Delta y_{ij}, \tag{10}$$

where

$$\Delta x_{ij} = (x_{i,1} - x_{j,1}) \text{ and } (x_{i,2} - x_{j,2}) \text{ and} \ldots$$
$$\Delta y_{ij} = y_i - y_j . \tag{11}$$

Here the negative sign has a special meaning, different from the conventional meaning of negation in fuzzy mathematics.

Each of the clustering regions $\mathbf{C}^{11}$, $\mathbf{C}^{12}$, $\mathbf{C}^{21}$ and $\mathbf{C}^{22}$ constitutes a specialized expert base. The specialization is given by the fuzzy sets $\mathbf{C}_1^1$, $\mathbf{C}_1^2$, $\mathbf{C}_2^1$ and $\mathbf{C}_2^2$.

Now, asking the clustering region $\mathbf{C}^{11}$ of the expert base a question $\mathbf{A}_Q^{11}$, we obtain a reply $\mathbf{B}_r^{11}$. This also applies to the clustering region $\mathbf{C}^{12}$ and the question $\mathbf{A}_Q^{12}$. The difference between the questions $\mathbf{A}_Q^{11}$ and $\mathbf{A}_Q^{12}$ si referred to as the delta difference of the questions $\mathbf{A}_Q^{11-12}$.

This question $\mathbf{A}_Q^{11-12}$ can be put to the delta clustering region $\mathbf{C}^{11-12}$. This delta clustering region is formed on the basis of application. In the following, we will always give preference to such treatment where no sign ambiguity occurs. The whole procedure can be written formally by means of the relations

$$\mathbf{A}_Q^{11} \rightarrow \mathbf{C}^{11} \rightarrow \mathbf{B}_R^{11}$$
$$\mathbf{A}_Q^{12} \rightarrow \mathbf{C}^{12} \rightarrow \mathbf{B}_R^{12}$$
$$\mathbf{A}_Q^{11} - \mathbf{A}_Q^{12} = \Delta\mathbf{A}_Q^{11-12} \tag{12}$$
$$\Delta\mathbf{A}_Q^{11-12} \rightarrow \Delta\mathbf{C}^{11-12} \rightarrow \Delta\mathbf{B}_R^{11-12},$$

It is also very simple to check the calculation using the relation

$$\mathbf{B}_R^{11} + \Delta\mathbf{B}_R^{11-12} = \mathbf{B}_R^{12} . \tag{13}$$

Data transfer is always based on heuristics, since exact data transfer is impossible. If the behaviour in the region $\mathbf{C}^{22}$ could be directly predicted, data transfer would be unnecessary.

For the beginning, a transformation between delta models will be demonstrated on a simple example. Assume that the identity

$$\Delta \boldsymbol{C}^{11-12} = \Delta \boldsymbol{C}^{21-22} \qquad (14)$$

holds between the delta models. This identity expresses the fact that the changes from the fuzzy set $\boldsymbol{C}_2^1$ to $\boldsymbol{C}_2^2$ are independent of the absolute value of the fuzzy set in the second dimension. This implies that

$$\begin{aligned} &\Delta \boldsymbol{A}_Q^{11-12} \to \Delta \boldsymbol{C}^{11-12} \to \Delta \boldsymbol{B}_R^{11-12} \\ &\Delta \boldsymbol{A}_Q^{21-22} \to \Delta \boldsymbol{C}^{21-22} \to \Delta \boldsymbol{B}_R^{21-22} \\ &\Delta \boldsymbol{B}_R^{11-12} \doteq \Delta \boldsymbol{B}_R^{21-22} . \end{aligned} \qquad (15)$$

The last statement expresses the fact that the impossibility of setting up a delta model $\Delta \boldsymbol{C}^{21-22}$ is circumvented by the assumption of identity. Therefore, it is convenient to introduce a correction factor $\boldsymbol{k}$, the meaning of which is apparent from the relation

$$\Delta \boldsymbol{B}_R^{21-22} = \boldsymbol{k} \,.\, \Delta \boldsymbol{B}_R^{11-12} . \qquad (16)$$

To avoid problems associated with the fuzzy set multiplication, we will only deal with the numerical representations of the fuzzy sets $\boldsymbol{k}$ and $\Delta \boldsymbol{B}^{11-12}$. The product of fuzzy sets thus transforms into a product of two real numbers. The correction factor is generally dependent on all fuzzy sets constituting the clustering regions $\boldsymbol{C}^{11}$, $\boldsymbol{C}^{12}$ and $\boldsymbol{C}^{21}$.

An alternative approach to the data transformation consists in the direct transformation

$$\boldsymbol{B}_R^{22} = \boldsymbol{D} \times \boldsymbol{B}_R^{11} . \qquad (17)$$

The transformation factor $\boldsymbol{D}$, however, is more difficult to estimate because the absolute value $\boldsymbol{B}_R^{11}$ is transformed into the absolute value of the reply. Using the correction factor $\boldsymbol{k}$, only the correction of the change is estimated, which is relatively easier.

Case Study

The test file is taken from ref.[9], where it is given in the expert base form. We modified this base to suit our purpose.

Variables that come into account when assessing accidents are summarized in Table I. Available information on individual accidents along with the effort made to keep the dimension of the problem within acceptable limits led to the selection of the variables given below.

The expert base was set up based on expert estimates and actual accident records. This approach proved fully suitable; it enables errors in expert estimates to be located and gross errors in the primary data to be disclosed by deduction using cross-checking.

The following labelling is henceforth used:

PO	population density	LO	low
TT	terrain type	ME	medium
SW	speed of wind ($m s^{-1}$)	HG	high
TE	performance temperature (°C)	VF	very flat
PR	performance pressure (kPa)	FL	flat
MA	amount of material (kg)	NO	normal
PL	property loss (million dollars)	HI	hilly
PA	causalties	VH	very hilly

TABLE I

Variables related to the evaluation of repercussions of accidents

Factor	Independent variable	Dependent variable
Population	population density distance from the population centre	
Climatic conditions	predominating wind direction	
Performance	nature of process performance temperature properties of material quantity of material structure	
Location	distance from other hazardous equipment	
Control and safety system	specification of safety system and alarm	
Repercussions		proper losses causalities

TABLE II

Records of 37 accidents in the expert base form

No.	Independent variables						Dependent variables		Weight of statement
	PO	TT	SW	TE	PR	MA	PL	FA	
1	LO	VF	1	8	8	1	1		1
2	LO	VF	2	3	3	—	2		1
3	LO	VF	3	2	2	3	3		1
4	ME	VF	4	2	2	4	4		1
5	ME	VF	5	3	3	—	5		1
6	ME	VF	6	4	4	6	6		1
7	LO	VF	—	3	3	7	7		1
8	ME	VF	8	1	1	8	8		1
9	LO	VF	9	1	1	9	9		1
10	ME	VF	10	5	5	—	10		1
11	LO	VF	11	1	1	11	11		1
12	LO	NO	12	1	1	12	12		1
13	LO	FL	13	2	2	13	13		1
14	HG	VF	14	5	5	14	14		1
15	ME	VF	15	3	3	—	15		1
16	LO	VF	16	6	6	16	16		1
17	LO	VF	17	7	7	17	17		1
18	LO	—	—	8	8	18	18		1
19	LO	VF	19	2	2	19	19		1
20	LO	VF	20	9	9	20	20		1
21	LO	VF	21	3	3	21	21		1
22	HG	VF	22	1	1	—	22		1
23	LO	VF	23	10	10	23	23		1
24	LO	FL	24	11	11	—	24		1
25	LO	VF	25	1	1	—	25		1
26	LO	VF	26	3	3	—	26		1
27	HG	HI	27	1	1	—	27		1
28	LO	NO	—	1	1	—	28		1
29	LO	VF	29	1	1	29	29		1
30	HG	VF	30	1	1	—	30		1
31	LO	VF	31	6	6	31	31		1
32	HG	NO	32	1	1	32	32		1
33	ME	VF	33	1	1	—	33		1
34	LO	VF	34	1	1	—	34		1
35	LO	VF	35	2	2	35	35		1
36	—	—	—	2	2	36	36		1
37	ME	VF	37	1	1	37	37		1

An expert base record comprising 37 accidents ($n = 37$ in relation (*1*)) is given in Table II. Table III specifies fuzzy sets used for the representation of the expert estimates (heuristics). The form of a fuzzy set is given by the constants a, b, c, d (Fig. 1). Table IV summarizes those variables that are considered deterministic (non-fuzzy). The clustering regions $\mathbf{C}^{11}$, $\mathbf{C}^{12}$ and $\mathbf{C}^{21}$ (see Eqs (*9*) and Fig. 2) are defined in Table VI.

TABLE III

Linguistic variables, their values and transformation to fuzzy variables

Linguistic variable	Linguistic value	*a*	*b*	*c*	*d*
PO	LO	0	0	100 000	102 000
	ME	100 000	102 000	1 000 000	1 020 000
	HG	1 000 000	1 020 000	5 000 000	5 000 000
TT	VF	0	0	91	96
	FL	91	96	183	192
	NO	183	192	457	480
	HI	457	480	914	960
	VH	914	960	1 829	1 829
TE	1	343	350	370	377
	2	60	260	280	300
	3	−105	−103	−48	−45
	4	150	370	400	400
	5	40	40	75	80
	6	50	80	100	102
	7	40	500	600	600
	8	161	165	175	179
	9	245	250	300	306
	10	833	850	860	877
	11	224	230	250	255
PR	1	95	101	102	106
	2	1 000	200 000	260 000	300 000
	3	60	66	80	90
	4	95	200	500	560
	5	360	370	490	500
	6	506	1 519	3 039	3 100
	7	400	2 500	3 500	3 600
	8	1 489	1 520	1 520	1 550
	9	95	101	102	106
	10	95	101	102	106
	11	6 949	7 091	7 091	7 233

Table IV
Values of the variables SW, MA, PL and FA in accident records

No.	Independent variables		Dependent variables	
	SW	MA	PL	FA
1	5·0	18 000	2·4	1
2	3·0	1	3·3	0
3	5·0	180	10·9	2
4	3·8	2 500	89·2	6
5	3·0	1	10·8	0
6	3·0	19 300	3·0	0
7	1·0	10 000	15·7	3
8	3·0	9 000	111·7	7
9	6·8	55 000	86·3	2
10	5·0	1	23·6	3
11	3·0	23 000	10·3	0
12	5·0	114 000	83·3	0
13	5·0	450	16·3	3
14	5·0	12 000	0·4	0
15	3·5	1	38·8	1
16	0·5	4 000	26·4	4
17	4·4	4 200	26·4	1
18	1·0	36 000	140·5	28
19	3·0	900	26·4	0
20	5·0	7 600	26·5	2
21	5·3	5 450	70·6	14
22	2·3	1	1·8	0
23	3·0	300	4·8	0
24	5·0	1	21·7	1
25	5·0	1	26·4	3
26	1·0	1	44·1	0
27	4·0	1	32·1	3
28	1·0	1	18·6	3
29	5·0	3 100	28·8	0
30	6·0	1	59·7	3
31	4·0	12 700	53·7	5
32	5·0	1 600	26·4	5
33	3·0	1	5·5	2
34	1·8	1	26·4	2
35	5·0	680	0·6	0
36	1·0	10 000	0·8	0
37	3·0	10 000	0·8	0

TABLE V

Questions to clustering regions C^{11}, C^{12}, C^{21} and replies

Variable	a	b	c	d
		Question A_Q^{11}		
PO	50 000·0	60 000·0	60 000·0	70 000·0
TT	70·0	80·0	80·0	90·0
SW	4·0	5·0	5·0	6·0
TE	80·0	90·0	90·0	100·0
PR	1 600·0	1 650·0	1 650·0	1 750·0
MA	4 000·0	4 500·0	5 000·0	5 500·0
		Reply B_R^{11}		
Active statements				
3	25·08/0·0	25·7/0·44	27·13/0·44	27·12/0·0
4	25·08/0·0	25·2/0·13	27·52/0·13	27·20/0·0
		Question A_Q^{12}		
PO	99 000·0	100 000·0	100 000·0	101 000·0
TT	90·0	100·0	100·0	110·0
SW	4·0	5·0	5·0	6·0
TE	250·0	260·0	270·0	280·0
PR	270 000·0	270 000·0	270 500·3	271 000·0
MA	150·0	160·0	160·0	170·0
		Reply B_R^{12}		
Active statement				
1	10·36/0·0	10·51/0·296	11·3/0·296	11·45/0·0
		Question A_Q^{21}		
PO	450 000·0	450 500·0	450 500·0	451 000·0
TT	50·0	55·0	55·0	60·0
SW	4·0	5·0	5·0	6·0
TE	60·0	65·0	65·0	70·0
PR	410·0	430·0	430·0	450·0
MA	11 000·0	11 500·0	11 500·0	12 000·0
		Reply B_R^{21}		
Active statement				
4	22·4/0·0	23·33/0·77	23·87/0·77	24·78/0·0

TABLE VI
Clustering regions $\mathbf{C}^{11}$, $\mathbf{C}^{12}$ and $\mathbf{C}^{21}$

No.	Independent variables						Dependent variable
	PO	TT	SW	TE	PR	MA	PL
				Region $\mathbf{C}^{11}$			
1	ME	VF	8	1	1	8	8
2	HG	VF	14	5	5	14	14
3	LO	VF	16	6	6	16	16
4	LO	VF	17	7	7	17	17
5	LO	VF	20	9	9	20	20
6	LO	VF	23	10	10	23	23
7	LO	VF	29	1	1	29	29
8	LO	VF	31	6	6	31	31
9	HG	NO	32	1	1	32	32
10	ME	VF	37	1	1	37	37
				Region $\mathbf{C}^{12}$			
1	LO	VF	3	2	2	3	3
2	ME	VF	4	2	2	4	4
3	LO	VF	7	3	3	7	7
4	LO	FL	13	2	2	13	13
5	LO	VF	16	6	6	16	16
6	LO	VF	17	7	7	17	17
7	LO	VF	19	2	2	19	19
8	LO	VF	21	3	3	21	21
9	LO	VF	35	2	2	35	35
10	—	—	—	2	2	36	36
				Region $\mathbf{C}^{21}$			
1	LO	VF	1	8	8	1	1
2	ME	VF	6	4	4	6	6
3	LO	VF	9	1	1	9	9
4	ME	VF	10	5	5	10	10
5	LO	VF	11	1	1	11	11
6	LO	NO	12	1	1	12	12
7	LO	18	18	8	8	18	18
8	HG	VF	22	1	1	22	22
9	LO	VF	25	1	1	25	25
10	HG	HI	27	1	1	27	27
11	LO	NO	28	1	1	28	28
12	HG	VH	30	1	1	30	30
13	LO	VF	31	6	6	31	31
14	ME	VF	33	1	1	33	33
15	LO	VF	34	1	1	34	34

By putting the clustering regions in Table VI questions which are collected in Table V, the statements given in Table V are activated. The serial numbers in Table V apply within the corresponding clustering regions only; the correspondence of these

TABLE VII

Model ($\Delta \mathbf{A}_Q^{11-12} \rightarrow \Delta \mathbf{C}^{11-12} \rightarrow \Delta \mathbf{B}_R^{11-12}$). Clustering region $\Delta \mathbf{C}^{11-12}$

No.	Independent variables												Dependent variables	
	PO1	PO2	TT1	TT2	SW1	SW2	TE1	TE2	PR1	PR2	MA1	MA2	PL1	PL2
1	ME	ME	VF	VF	8	4	1	2	1	2	8	4	8	4
2	LO	LO	VF	VF	16	17	6	7	6	7	16	17	16	17
3	LO	LO	VF	VF	20	16	9	6	9	6	20	16	20	16
4	LO	LO	VF	VF	29	16	1	6	1	6	29	16	29	16
5	HG	LO	NO	VF	32	16	1	6	1	6	32	16	32	16
6	LO	LO	VF	VF	16	3	6	2	6	2	16	3	16	3
7	LO	LO	VF	VF	17	3	7	2	7	2	17	3	17	3
8	LO	LO	VF	VF	29	7	5	1	5	1	29	7	29	7
9	LO	LO	VF	VF	20	7	9	3	9	3	20	7	20	7
10	LO	LO	VF	VF	23	3	10	2	10	2	23	3	23	3
11	LO	LO	VF	VF	17	16	7	6	7	6	17	16	17	16

Question $\Delta \mathbf{A}_Q^{11-12}$

PO1:	50 000·0	60 000·0	60 000·0	70 000·0
PO2:	99 000·0	100 000·0	100 000·0	101 000·0
TT1:	70·0	80·0	80·0	90·0
TT2:	90·0	100·0	100·0	110·0
SW1:	4·0	5·0	5·0	6·0
SW2:	4·0	5·0	5·0	6·0
TE1:	80·0	90·0	90·0	100·0
TE2:	250·0	260·0	270·0	280·0
PR1:	1 600·0	1 650·0	1 650·0	1 750·0
PR2:	270 000·0	270 500·0	270 500·0	271 000.0
MA1:	4 000·0	4 500·0	5 000·0	5 500·0
MA2:	150·0	160·0	160·0	170·0

Reply $\Delta \mathbf{B}_R^{11-12}$

Active statements	*a*	*b*	*c*	*d*
6	14·44/0·0	14·74/0·286	16·05/0·286	16·27/0·0
7	14·72/0·0	14·82/0·128	16·28/0·128	16·30/0·0

numbers and the serial numbers in Table II is clear from the values of the dependent variable PL.

The $\Delta\mathbf{C}^{11-12}$ model is specified in Table VII, the correction factors $\boldsymbol{k}$ are given in Table VIII, the direct correction factor $\boldsymbol{D}$ is specified in Table IX. To enable the precision of the results obtained to be examined, the model of the clustering region

TABLE VIII

Correction factor $\boldsymbol{k}$

Independent variables	$\boldsymbol{k}$			
PO1	50 000·0	60 000·0	60 000·0	70 000·0
PO2	99 000·0	100 000·0	100 000·0	110 000·0
PO3	450 000·0	450 000·0	450 500·0	451 000·0
PO4	34 000·0	34 500·0	34 500·0	34 800·0
TT1	70·0	80·0	80·0	90·0
TT2	90·0	100·0	100·0	110·0
TT3	50·0	55·0	55·0	60·0
TT4	50·0	60·0	60·0	70·0
SW1	4·0	5·0	5·0	6·0
SW2	4·0	5·0	5·0	6·0
SW3	4·0	5·0	5·0	6·0
SW4	2·5	2·7	2·7	3·2
TE1	80·0	90·0	90·0	100·0
TE2	250·0	260·0	270·0	280·0
TE3	60·0	65·0	65·0	70·0
TE4	−90·0	−80·0	−80·0	−70·0
PR1	1 600·0	1 650·0	1 650·0	1 750·0
PR2	270 000·0	270 000·0	270 000·0	271 000·0
PR3	410·0	430·0	430·0	450·0
PR4	70·0	80·0	80·0	90·0
MA1	4 000·0	4 500·0	5 000·0	5 500·0
MA2	150·0	160·0	160·0	170·0
MA3	11 000·0	11 500·0	11 500·0	12 000·0
MA4	13 800·0	13 900·0	13 900·0	14 000·0
		Reply		
Active statement	*a*	*b*	*c*	*d*
6	0·1/0·0	0·186/0·286	0·714/0·286	0·76/0·0

$\boldsymbol{C}^{22}$ is given in Table X. By means of the clustering region $\boldsymbol{C}^{21}$, an estimate of the loss of approximately 23·60 million dollars is obtained in reply to the question $\boldsymbol{A}_Q^{21}$.

By the delta cluster $\Delta\boldsymbol{C}^{11-12}$ the delta loss $\Delta\boldsymbol{B}^{11-12}$ is estimated to 15·40 million dollars. By using the correction factor $\boldsymbol{k}$ from Table IX (in its numerical representation), the delta loss $\Delta\boldsymbol{B}^{21-22}$ is determined from the relation

$$\Delta\boldsymbol{B}_R^{21-22} = \boldsymbol{k} \,.\, \Delta\boldsymbol{B}_R^{11-12} = 0{\cdot}45 \times 15{\cdot}40 \,.\, 10^6 = 6{\cdot}93 \,.\, 10^6 \,.$$

Thus, the relation

$$\boldsymbol{B}_R^{22} = \boldsymbol{B}_R^{21} - \Delta\boldsymbol{B}_R^{11-12} = (23{\cdot}60 - 6{\cdot}93) \,.\, 10^6 = 16{\cdot}67 \,.\, 10^6 \text{ dollars}$$

is used for estimating the losses. Using the direct correction factor $\boldsymbol{D}$ in Table IX the losses are estimated to

$$\mathrm{B}_R^{22} = \boldsymbol{D} \times \boldsymbol{B}_R^{11} = (26{\cdot}40 \times 0{\cdot}46) \,.\, 10^6 = 12{\cdot}16 \,.\, 10^6 \text{ dollars} \,.$$

TABLE IX

Direct correction **D**

Independent variables	D			
PO1	50 000·0	60 000·0	60 000·0	70 000·0
PO2	34 000·0	34 500·0	34 500·0	34 800·0
TT1	70·0	80·0	80·0	90·0
TT2	50·0	60·0	60·0	70·0
SW1	4·0	5·0	5·0	6·0
SW2	2·5	2·7	2·7	3·2
TE1	80·0	90·0	90·0	100·0
TE2	−90·0	−80·0	−80·0	−70·0
PR1	1 600·0	1 650·0	1 650·0	1 750·0
PR2	70·0	80·0	80·0	90·0
MA1	4 000·0	4 500·0	5 000·0	5 500·0
MA2	13 800·0	13 900·0	13 900·0	14 000·0
	Resultant direct correction factor **D**			
Active statements	*a*	*b*	*c*	*d*
9	0·1/0·0	0·233/0·44	0·688/0·44	0·76/0·0
10	0·1/0·0	0·138/0·128	0·739/0·128	0·76/0·0

For a check, we used a direct calculation by means of the cluster $\mathbf{C}^{22}$ (which is usually empty, whereby the requirement of data transfer arises): $\mathbf{B}_R^{22} = 13\cdot30 \,.\, 10^6$ dollars.

The solution of the above case is based on real data. More extensive data, with respect to the number of accidents and/or the number of parameters, are largely unavailable.

The next case study will concern nuclear power station steam generators. Primary data were extracted from ref.[10] The total number of steam generators included was 110, the total number of parameters was 17. The latter were divided into the following groups: design and technology, time, material, water pretreatment and reliability.

The first group comprises the basic design quantities (e.g. the heat exchange area) and production technology data. Time parameters include the time the commercial operation was started and time data concerning the water pretreatment methods. Information from the materials group only comprises data of the structural materials

TABLE X

Check calculation $\mathbf{A}_Q^{22} \rightarrow \mathbf{C}^{22} \rightarrow \mathbf{B}_R^{22}$. Clustering region $\mathbf{C}^{22}$

No.	Independent variables						Dependent variabe
	PO	TT	SW	TE	PR	MA	PL
1	LO	VF	2	3	3	2	2
2	ME	VF	5	3	3	5	5
3	LO	VF	7	3	3	7	7
4	ME	VF	15	3	3	15	15
5	LO	FL	24	11	11	24	24
6	LO	VF	26	3	3	26	26
7	36	36	36	2	2	36	36

Question $\mathbf{A}_Q^{22}$

PO:	34 000·0	34 500·0	34 500·0	34 800·0
TT:	50·0	60·0	60·0	70·0
SM:	2·5	2·7	2·7	3·2
TE:	−90·0	−80·0	−80·0	−70·0
PR:	70·0	80·0	80·0	90·0
MA:	13 800·0	13 900·0	13 900·0	14 000·0

Reply $\mathbf{B}_R^{22}$

Active statement	*a*	*b*	*c*	*d*
1	12·64/0·0	13·05/0·625	13·55/0·625	13·9/0·0

(e.g. pipes). Reliability information is represented by the failure intensity and the total time of failure-free performance.

Details of the application of the data transfer are too extensive to be given in this paper. Practical experience, however, can be summarized briefly. A number of data (on materials for instance) are discrete by nature. Thereby, the initial expert base constituted by 110 statements ($n = 110$ in relation (*1*)) decomposes into a number of sub-bases since discrete data cannot be fuzzified.

Setting up clusters under these conditions, we find that only two steam generators belong to a common cluster; these are two identical steam generators used in one nuclear power station.

In this manner, 100 clusters are actually obtained, all of which – save one – contain a single steam generator each. This is an atypical situation where the indiviual clusters $\mathbf{C}^{ij}$ (see Eqs (*9*)) are constituted by a single statement each.

It appeared suitable to repeat the delta transfer several times. We can begin the dialogue, e.g., by transforming within the design-and-technology cluster. This cluster is obtained by eliminating from the expert base all parameters that do not concern design or technology. For instance, the water pretreatment data are completely ignored. In this manner we obtain specialized clusters, by means of which experience is transformed to the region of design and technological parameters we need. Thus, we gain data about those steam generators in the expert base that are the most fuzzy-similar to the design and technological parameters we are interested in. Next, we can go on to the clustering from the point of view of another set of parameters, e.g. from the reliability point of view.

The most fuzzy-similar steam generators obtained by clustering from a point of view are almost always dissimilar in clustering from the set of parameters used before. The only way out consists of a dialogue and compromise between the similarities emerging from the treatment from different aspects. It is here that expertise must be included whose application is made possible by human logic.

Trends of Potential Interest

A fundamental shortcoming of the data transfer as outlined in this paper is in the fact that deep knowledge cannot be handled in the present model. This even precludes utilization of universal laws such as, e.g., the law of conservation of mass.

Indirectly, these laws of nature can be respected by the fact that the statements of relation (*1*) are consistent with them. However, they become fuzzy in the process of transformation of knowledge. It would be therefore very useful, in the case of universally valid deep knowledge represented by laws of nature, to employ the results of the transformation for minimizing the necessary fuzzification.

From this viewpoint, integration of qualitative and quantitative knowledge is of great utility. Ref.[11] gives one of the possible routes where qualitative information can be presented as laws of nature. Qualitative information then is used to actually set up a qualitative model. In this way the maximum possible degree of exploitation of information is reached.

The problem of qualitative knowledge is highly topical with respect to reliability and safety. However, the subject is discrete in nature, and so the treatment will meet with difficulties of the same type as those encountered when setting up artificial intelligence algorithms.

Subjectivity has been studied theoretically, and attempts at practical applications exist[12]. As yet, the knowledge of subjectivity has only been employed to increase the objectivity of expert bases, but psychological knowledge can be now utilized for objectivization of the dialogue with the expert base[13].

With the use of cognitive algorithms, which initially were set up for sociological purposes, a question the expert system is asked by the user can be modified so that while the changes in the question are minimal, the clarity of the reply can improve, sometimes considerably[14].

Another apparent trend consists not in bringing in new methods but in systemizing all the knowledge available. Studying accidents in chemical industry, we face an interdisciplinary problem. Let us name at random some of the relevant branches of science and technology involved in the assessment of accidents[15–18]: chemical engineering, properties of substances, physical chemistry, quantum mechanics, technology, materials science, control and regulation, mathematics, statistics, etc. In all of them, knowledge has to be fixed in an appropriate manner to make full use of it, and the optimum information and knowledge structures must be constructed. If at least semi-optimum information structures have been set up, their optimal synthesis will be made possible, facilitating the solution of the ultimate problem, e.g. detection of the primary failure.

Work on such information structures is in progress (e.g. ref.[19]). This is, naturally, very difficult and laborious. It is imperative that improper classification of pieces of knowledge into individual classes, which with the progress in computer science and knowledge engineering has lost its significance, be abandoned. Such pigeon--holing not only does not add to the quality of knowledge storage but it even can detract from the resultant quality of the information structure.

Knowledge related to chemical equipment failure should be therefore presented with all its original liaisons rather than through compartmentalized pieces of knowledge within, e.g., chemical engineering and technology. As a matter of fact, such division is never sharp and, actually, never was.

The algorithm suggested in this paper is not the only one possible; a number of variants exist. Although not described here, such variants are not difficult to derive from the basic idea, but in the present stage, none can be recommended. It is even hard to say under which conditions the clustering algorithm should be employed for data transfer in reliability as such. A number of problems remain to be solved. Still it can be claimed that:

primary data should be collected to the highest possible extent and stored in the form in which they are available – in the numerical, fuzzy or verbal form;

optimal knowledge structure should be set up so as not to force the user to distort the reliability just because of insufficient flexibility of the information structure;

expert teams as wide as possible should be invited to cooperate;

data acquisition should be automated, e.g. by coupling the information acquisition system for control and regulation with that for reliability.

REFERENCES

1. Theofamons T. G.: Chem. Eng. Sci. *8*, 1615 (1983).
2. Wingender H. J. S.: *Reliability Data Collection and Use in Risk and Availability Assessment.* Springer, Berlin 1986.
3. Dohnal M.: Chem. Prum. *37*, 617 (1987).
4. Fargin R., Halpern J. Y., Vardi M. Y.: *Research Report RJ 6250.* IBM Research, San Jose 1988.
5. Jambu M., Lebeaux M. O.: *Cluster Analysis and Data Analysis.* North-Holland, Amsterdam 1983.
6. Dubois D., Prade H.: *Fuzzy Sets and Systems.* Academic Press, New York 1980.
7. Vaija P., Turunen I., Järvelainen M., Dohnal M.: Microelectron. Reliab. *25*, 369 (1985).
8. Dohnal M.: Chem. Eng. J. *30*, 71 (1985).
9. Vaija P., Järvelainen M., Dohnal M.: *Symp. Inst. Chem. Eng., EPCE Event No. 322, Manchester 1985*; p. 397.
10. Tatone O. S., Pathania R. S.: Nucl. Safety *25*, 373 (1984).
11. Falkenhainer B. C., Michalski R.: Machine Learning *1*, 361 (1986).
12. Olson J. R., Rueter H.: *Tech. Report No. 10,* University of Michigan, Ann Arbor 1987.
13. Novick L. A.: J. Exp. Psychol. *14*, 510 (1988).
14. Dohnal M.: *Chemdata 1988, Proc. Use of Computers in Chem. Eng.*, Vol 1, p. 193. Gothenburg 1988.
15. Steinberg D. M., Hunter W. G.: Technometrics *26*, 71 (1984).
16. Nachtsheim C. J.: J. Quality Technol. *19*, 132 (1987).
17. Rosen E. M., Adams R. N.: Comput. Chem. Eng. *11*, 720 (1987).
18. Bainbridge L.: Future Comput. Systems *1*, 149 (1986).
19. Bowman M. D., Nordmark G. E., Yao J. P. T.: Int. J. Approx. Reasoning *1*, 197 (1987).

Translated by P. Adámek.