FUZZY BASED EXPERT SYSTEM FOR ANALYSIS OF ACCIDENTS

P. Vaija[+], M. Järveläinen[+], M. Dohnal[++]

Basic principles in creating fuzzy based expert systems for problems dealing with subjective, ill-defined and uncertain knowledge are discussed. These principles are applied in a simple test system which is built up for accident analysis. The test system is based on an actual record set of former accidents.

INTRODUCTION

In the process industries, analyses of former accidents are needed to design safer processes with lower risks and to organize protection. Also legistative authorities and insurance companies may need information about risks in the processes /9/.

For present needs, there are some data banks storing reports of accidents /8/. Some companies and organisations also have their own investigations concerning internal accidents. However, information is in many cases limited to notes made at the time and to eyewitness reports. Accordingly, the knowledge in data banks can be very poor and unreliable. Data from one particular process may also be unsuitable for predicting the possibility of an accident or its consequences in a different process.

In this paper, as far as we know the principles to be applied in developing an expert system for accident analysis are presented first time. The expert system can widen the use of these data banks and increase the forecasting power of the system. Expert systems based on fuzzy simulation are advantageous in this field because, as mentioned, reports about accidents are often subjective and uncertain. Fuzzy mathematics can deal with such data and put them into a form acceptable to the computer. Fuzzy simulation also brings the models close to human reasoning. The basic knowledge of fuzzy mathematics is presented in /20/. Fuzzy applications to chemical engineering are presented in /15,16,18/.

+ Laboratory of Chem. Eng., Helsinki University of Technology, Finland
++ Department of Chem. Eng., Technical University of Brno. Czechoslovakia

EXPERT SYSTEMS

Expert systems are information systems which can give answers to the users questions dealing with the special knowledge in their expert base.

Knowledge in the expert base is usually formulated in rules determining the conditions under which the rules are valid and the consequences of these conditions. Knowledge in these rules can include data from laboratory or pilot plant experiments, measuring data from processes, data from literature etc.

However, the backbone of expert systems is formed by rules based on experts reasoning and knowledge of causal connections between conditions and consequences of the problem under study. This reasoning includes typically subjective and inexact conclusions, too /19/.

The second part of an expert system is an intelligent interface. When a question is passed to the expert system, this subsystem tries to answer on the basis of the knowledge in the expert base. In some expert systems there is a reasoning algorithm in the intelligent interface. This helps the system itself during the problem solving by deciding or planning which is the next step in evaluation.

In this approach the expert base is adjusted to the universal expert system SENECA /11/ which is based on fuzzy simulation. The evaluating algorithm used is the program system CONFUCIUS /12,13/. The advantage of this system is that it can accommodate rather heterogenous and partially inconsistent data of different ranges of accuracy.

The expert base is developed on the basis of fuzzy reasoning. The knowledge in the expert base is given in the form of fuzzy conditional statements

$$\text{if } A_i \text{ then } B_i \tag{1}$$

where A_i is the n-dimensional set of fuzzy values of independent variables. B_i is the corresponding value of dependent variable. Values of variables are given deterministically or in a fuzzy way depending on their accuracy and specified by their membership functions. In the set of conditional statements knowledge with more exact data can be stressed by higher weight factors ($0< w \leq 1$).

When the user poses a question to this expert system the fuzzy evaluation program (CONFUCIUS) gives answers using the fuzzy expert base in the form of conditional statements (1). If the evaluation algorithm cannot find an answer, e.g. if the question is too far from the space determined by the conditional statements, the reasoning algorithm begins to fuzzify the question. The simplest suitable algorithm is based on the gradual fuzzification of the question. However, it is also possible to use a more sophisticated algorithm which takes into account the whole contents of the expert base and different weight factors defined for these variables and statements.

FUZZY BASED ACCIDENT ANALYSIS

In order to develop an expert base in the fuzzy expert system SENECA data of accidents are first transformed into the form of fuzzy conditional statements (1). Next, general rules specifying the conditions under which accidents might happen and their possible consequences ought to be considered and transformed also into the form of fuzzy conditional statements. This consideration can be made intuitively or on the basis of an accident record set, if available. The incorporation of the accident record set to the expert base increases the reliability of the analysis and prediction.

One proposal of variables which could specify the situation during accidents and their consequences is given in Table 1. Here the values of independent variables of a certain accident specify the n-dimensional fuzzy set A_i and the values of consequences B_i in the formula (1). Thus for each accident one conditional statement is written. Because in the first record set knowledge is more detailed variables are fed into the system deterministically. The second set indicates general rules for accident so variables in this set are given in a fuzzy way.

If one wants to analyze and predict several consequences of the same accident, then for every consequence analysis its own expert base should be developed. The independent variables are naturally the same, only the dependent variable is different.

The expert system can be exploited in two ways. First, consequences of an accident not included in the expert base can be predicted. Second, in the design stage, consequences of a hypothetical accident can be fixed and the expert base asked to give maximum values of one or some independent variable(s) under study until the dependent variable reaches the prescribed value.

TEST EXAMPLE

In this study the idea of developing an expert base for analysing accidents was tested by an example. Two simple fuzzy expert bases were built, one for predicting property losses (PROEXP) and the other fatalities (FATEXP) on the basis of an accident record set. This was found from the literature /14/.

From this set eleven different processes were chosen dealing with petrochemical industry and refineries. The total number of accidents in the set was 37.

From primary data /14/ and literature the following was found to specify roughly the situation during an accident:

PO = population near the plant /5,6/
TT = altitude of terrain /5,6/
SW = speed of the prevailing wind /7/

The type of the process was specified by the running temperature /1-4,17/ and pressure /1-4,17/. The amount of the material available was found from /14/.

The quality of the material is not included in the list of independent variables because all materials except vinyl chloride were merely qualified inflammable. Vinyl chloride is toxic in addition to that.

In literature /14/ property losses and fatalities were given as consequences.

The first step in developing the fuzzy expert base is to specify deterministic or linguistic values for variables. As linguistic values expressions like low, medium and high can be used. Usually at least three linguistic values are needed for every variable. Next these values are transformed into fuzzy sets by specifying their membership functions. For practical purposes the grade of membership function is specified by four points /see Fig. 2/.

The basic consept of fuzzy mathematics states that an element can belong partly to a set. If a conventional set is denoted by

$$U = \{ x \}$$

then a fuzzy set A in the universe U is a set of ordered pairs

$$A = \{(x, m_A(x)) \} \tag{2}$$

where $m_A(x)$ = grade of membership of x in A

$$0 \leq m_A(x) \leq 1$$

For an element of a conventional set it is possible either to belong totally to a set or not to belong to this set

$$\begin{aligned} m_A(x) &= 1, \quad x \in A \\ m_A(x) &= 0, \quad x \notin A \end{aligned} \tag{3}$$

Let us suppose that the grade of membership function in Fig 2 presents the normal temperature of the process. Then in the area bc the temperature belongs totally to the fuzzy set normal temperature i.e. the grade of membership is one. In the areas ab and cd the temperature belongs to the fuzzy set normal temperature with a certain grade of membership.

The form of the grade of membership function depends on the accuracy of the knowledge. In practice the areas ab and cd in Fig. 2 can mean the inaccuracy in data, for example the measuring accuracy. If the variable has an exact value then in the grade of membership function b=c. Deterministic values are given as a point i.e. a=b=c=d /see Fig. 1/.

When developing the accident record set deterministic values were found for the speed of prevailing wind, amount of material available, property losses and fatalities / see Table 2/. Other data were so inexact that linguistic values were preferred /see Table 3/.

Because there were eleven different process types, as many linguistic values were chosen for process temperature and pressure. Those linguistic values were presented using numbers for the sake of convenience. If there were different types of the same process then the linguistic values were accentuated to the most dangerous areas. Linguistic values and their membership functions are shown in Table 4.

After this the relationships between circumstances and consequences were specified. In fuzzy based expert systems these relationships are given by a multidimensional set of fuzzy conditional statements (1)

$$\begin{aligned} &\text{if } A_{1,1} \text{ and } A_{1,2} \; . \; . \; . \text{ and } A_{1,n} \text{ then } B_1 \text{ or} \\ &\text{if } A_{2,1} \text{ and } A_{2,2} \; . \; . \; . \text{ and } A_{2,n} \text{ then } B_2 \text{ or} \\ &. \\ &. \\ &. \\ &\text{if } A_{m,1} \text{ and } A_{m,2} \; . \; . \; . \text{ and } A_{m,n} \text{ then } B_m \end{aligned} \tag{4}$$

Here the fuzzy sets $A_{i,j}$ mean circumstances and the sets B_i corresponding consequences.

For every accident one conditional statement is written. In Table 5 there is a set of conditional statements. If there was no information available of independent variables there is a bar (-) in the statement. Then the corresponding values in Tables 2 and 4 are meaningless, usually the value in these tables is one. If the values of dependent variables were not found average values were used.

The general rules for expert bases were formulated after limiting the location of hazardous processes to areas where population is under 100 000 persons and where the slope of terrain is low.

Linguistic values and their membership functions were specified to those variables which were determined deterministically in the first record sets. These linguistic values and their membership functions are shown in Table 6.

The general conditional statements were written applying the same principle as before. In Table 7 there are some examples of these statements.

The expert bases PROEXP and FATEXP were built up by connecting the record set to the corresponding speculation based set. The number of the statements in both expert bases were 108.
Because the reliability of the knowledge in these two sets were different, weight factors were given to the conditional statements. The weight factor of the first sets were the highest possible i.e. one because of exact knowledge. The weight factor of the second sets were lower. The choice of the weight factors is subjective. In this approach the weight factor of the second sets were 0.80.

The expert systems were tested by five questions presented in Table 8. The evaluating algorithm tested the values of the question against the knowledge in the form of fuzzy conditional statements in the expert bases. The fuzzy results were transformed to numerical ones by calculating the centers of gravities for the independent variables by the following equation

$$C = \frac{\sum_{i=1}^{r} x_i \ m_i}{\sum_{i=1}^{r} m_i} \qquad (5)$$

where r is the cardinality of a universe, x_i is an element of this universe with the grade of membership m_i /21/.

Answers are shown in Table 9. The fuzzy evaluation system could not find answers to the question 2 and 5. Therefore, values of variables were fuzzified gradually as shown in Fig. 1 and 2. These fuzzified values of variables in question 2 and 5 are shown in Table 10. The amount of material was fuzzified more quickly because the results in record sets seemed to be more dependent on this variable. From Table 11 it can be seen that after the first fuzzification an answer was found to the second question. The ratio fatalities/property losses is higher than in other because this answer is based on the knowledge in the record sets. All the other questions have activated conditional statements from the speculation based sets. After three fuzzifications, an answer was not found to the question 5. Further fuzzification was not reasonable because then the question was considered to be too far from the original.

CONCLUSION

According to the results of test examples it seems that the fuzzy mathematics and simulation is an effective way to transform data of accident reports into a suitable form for building an expert base.

The fuzzy expert system for accident analysis is a useful method in predicting consequences of accidents not included in this expert base. The system can be used also in design purposes by calculating maximum values of some variables specifying the process or the circumstances corresponding to the prescribed values of consequences.

NOMENCLATURE

PO = population, persons
TT = altitude of terrain from sea level, m
SW = speed of the prevailing wind, m s^{-1}
TE = operating temperature, ^{o}C
PR = operating pressure, kPa
MA = amount of material, kg
PL = property loss, millions of dollars
FA = fatalities, person

REFERENCES

1. Komi, R. (1982) Öljystä muoveihin, Neste Oy, Helsinki.
2. Hydrocarbon Processing, 56, pp. 115-242.
3. Hydrocarbon Processing, 54, pp. 97-224 .
4. Shreve, R. N., Brink, J. A., (1977) Chemical Process Industries, 4. ed., McGraw-Hill Book Company, New York.
5. The New Oxford Atlas, Oxford University Press, 1978.
6. The Times Atlas of the World, Times Book, London, 1977.
7. World Survey of Climatology, vol. 5, 6, 8, 11, 12, Elsevier Publishing Company, Amsterdam, 1970.
8. Daniels, J. T., Holden, B. L. (1983) IChem E Symposius Series no 80, Harrogate, pp. G 33-45.
9. Opschhor, G., Schecker, H. G. (1983) IChem E Symposium Series no 80, Harrogate, pp. G 25-32.
10. Dudak. M., Kleinova, N., Mikula, O. (1984) Application of expert systems in chemical industry, 8. International CHISA Congress, Praque.
11. Dohnal, M., (1984) Application of universal expert systems in industry, Research report, TUB, Brno.
12. Dohnal, M. (1983) Computers in Industry, 4, 4, pp. 341-346.
13. Dohnal, M. (1984) Computers in Industry, 4, 4, pp. 347-352.
14. Davenport, J. A. (1983) IChem E Symposium Series no 80, Harrogate, pp. C 1-9.
15. Vaija, P., Turunen, I., Järveläinen, M., Dohnal, M. (1985), accepted to be published in Microelectronics and Reliability.
16. Dohnal, M., Turunen, I., Järveläinen, M. (1983) Kemia- Kemi, 10, 5, pp. 394-400.
17. Weissermehl, K., Arpe, H.-J. (1978) Industrial Organic Chemistry, Verlag Chemie, Weinheim.
18. Turunen, I. (1984) Acta Polytechnica Scandinavica, Chemical Technology and Metallurgy Series no 155, Helsinki.
19. Lehtinen, E. (1984) Tietotekniikkapäivät Otadata-83, Otapaino, Espoo, pp. B/1/1 - B/1/13.

20. Dubois, D., Prade. H. (1980) Mathematics in Science and Engineering no 144, Academic Press, New York.
21. Dohnal, M. (1984) Fuzzy simulation of chemical processes, Research report, TUB, Brno.

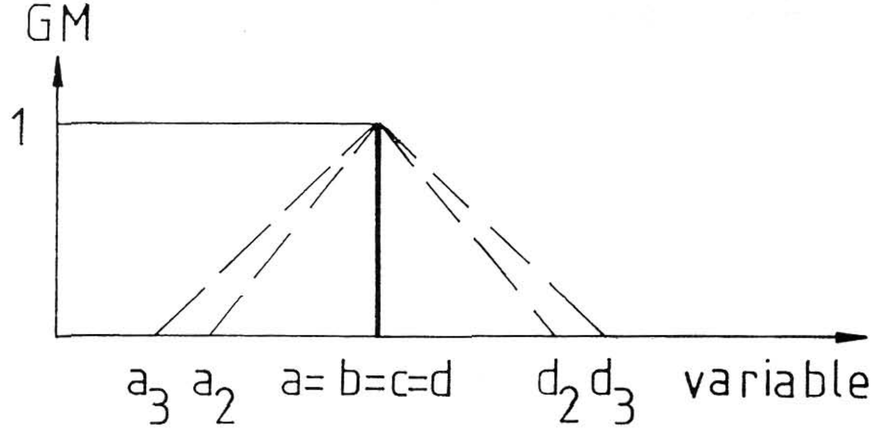

Figure 1. Gradual fuzzification of deterministic values. Fuzzification is marked by dashed lines.

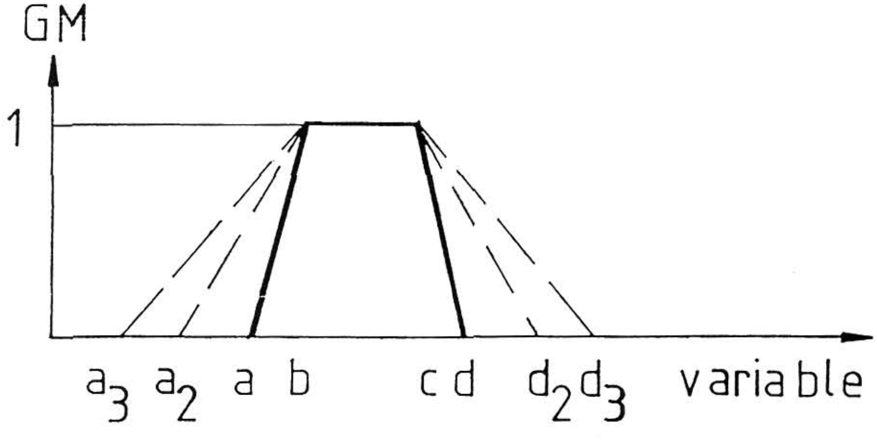

Figure 2. Gradual fuzzification of fuzzy values. Fuzzification is marked by dashed lines.

Table 1. Knowledge need for developing an expert system for accident analysis /15,16,18/.

section	independent variables	dependent variables
site selection	- density of population - distance from populated areas	
climate	- the prevailing wind	
process	- specification of process - operating temperature - operating pressure - quality of material - quantity of material - types of structures	
layout	- distances from other hazards and populated areas	
control and safety systems	- specifications of alarm and safety systems	
damage	- specification of primary reasons	
consequences		- property losses - fatalities

Table 2. Linguistic values of variables in record sets given as presented in Fig. 1.

sequence of accident	independent variables		dependent variables	
	SW	MA	PL	FA
1	5.0	18000	2.4	1
2	3.0	1	13.3	0
3	5.0	180	10.9	2
4	3.8	2500	89.2	6
5	3.0	1	10.8	0
6	3.0	19300	3.0	0
7	1.0	10000	15.7	3
8	3.0	9000	111.7	7
9	6.8	55000	86.3	2
10	5.0	1	23.6	3
11	3.0	23000	10.3	0
12	5.0	114000	83.3	0
13	5.0	450	16.3	3
14	5.0	12000	0.4	0
15	3.5	1	38.8	1
16	5.0	4000	26.4	4
17	4.4	4200	26.4	1
18	1.0	36000	140.5	28
19	3.0	900	26.4	0
20	5.0	7600	26.5	2
21	5.3	5450	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	3100	28.8	0
30	6.0	1	59.7	3
31	4.0	12700	53.7	5
32	5.0	1600	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	2500	26.4	0
37	3.0	10000	0.8	0

Table 3. Linguistic values and their membership functions of population and altitude of terrain given as presented in Fig. 2.

variable	linguistic value	membership function a	b	c	d
PO	LO	0	0	100000	102000
	ME	100000	102000	1000000	1020000
	HG	1000000	1020000	5000000	5000000
TT	VF	0	0	91	96
	FL	91	96	183	192
	NO	183	192	457	480
	HI	457	480	914	960
	VH	914	960	1829	1829

Table 4. Linguistic values of temperature and pressure in the processes given as presented in Fig. 2.

variable	linguistic value	membership function a	b	c	d
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
	1	95	101	102	106
	2	1000	200000	260000	300000
	3	60	66	80	90
	4	95	200	500	510
	5	360	370	490	500
	6	506	1519	3039	3100
	7	400	2500	3500	3600
	8	1489	1520	1520	1550
	9	95	101	102	106
	10	95	101	102	106
	11	6949	7091	7091	7233

Table 5. Conditional statements for record sets.

no	independent variables						dep. variable	weight factor	code in /14/
	PO	TT	SW	TE	PR	MA	PL or FA		
1	LO	VF	1	8	8	1	1	1	3
2	LO	VF	2	3	3	-	2	1	1
3	LO	VF	3	2	2	3	3	1	18
4	ME	VF	4	2	2	4	4	1	11
.	ME	VF	5	3	3	-	5	1	7
.	ME	VF	6	4	4	6	6	1	4
.	LO	VF	-	3	3	7	7	1	62
	ME	VF	8	1	1	8	8	1	6
	LO	VF	9	1	1	9	9	1	13
	ME	VF	10	5	5	-	10	1	102
	LO	VF	11	1	1	11	11	1	40
	LO	NO	12	1	1	12	12	1	77
	LO	FL	13	2	2	13	13	1	32
	HG	VF	14	5	5	14	14	1	22
	ME	VF	15	3	3	-	15	1	49
	LO	VF	16	6	6	16	16	1	105
	LO	VF	17	7	7	17	17	1	52
	LO	-	-	8	8	18	18	1	9
	LO	VF	19	2	2	19	19	1	57
	LO	VF	20	9	9	20	20	1	8
	LO	VF	21	3	3	21	21	1	60
	HG	VF	22	1	1	-	22	1	70
	LO	VF	23	10	10	23	23	1	74
	LO	FL	24	11	11	-	24	1	51
	LO	VF	25	1	1	-	25	1	136
	LO	VF	26	3	3	-	26	1	90
	HG	HI	27	1	1	-	27	1	94
	LO	NO	-	1	1	-	28	1	95
	LO	VF	29	1	1	29	29	1	81
	HG	VF	30	1	1	-	30	1	98
	LO	VF	31	6	6	31	31	1	99
	HG	NO	32	1	1	32	32	1	120
	ME	VF	33	1	1	-	33	1	142
	LO	VF	34	1	1	-	34	1	145
	LO	VF	35	2	2	35	35	1	10[illegible]
	-	-	-	2	2	36	36	1	33
37	ME	VF	37	1	1	37	37	1	143

Table 6. Linguistic values of variables in the speculation based sets given as presented in Fig. 2.

variable	linguistic value	a	membership function b	c	d
SW	LO	0.0	0.0	1.5	4.2
	HG	1.5	4.2	8.0	8.0
TE	LO	-150	0	70	140
	ME	70	140	350	700
	HG	350	700	1000	1000
PR	LO	0	100	500	2500
	ME	500	2500	15000	225000
	HG	15000	225000	300000	300000
MA	LO	0	0	13000	32000
	ME	13000	32000	65000	85000
	HG	65000	85000	115000	120000
PL	LO	0	0	20	26
	ME	20	26	72	100
	HG	72	100	140	150
FA	LO	0	0	2	3
	ME	2	3	10	11
	HG	10	11	29	30

Table 7. Conditional statements of speculation based sets.

no	PO	independent variables TT	SW	TE	PR	MA	dep. variable PL or FA	weight factor
1	LO	VF	HG	ME	LO	ME	ME	0.8
2	LO	VF	LO	ME	LO	LO	LO	0.8
3	LO	VF	LO	LO	ME	LO	LO	0.8
4	LO	VF	LO	HG	ME	ME	HG	0.8
5	LO	VF	LO	HG	LO	LO	ME	0.8
6	LO	FL	HG	LO	LO	ME	LO	0.8
7	LO	VF	HG	ME	HG	LO	ME	0.8
8	LO	FL	LO	HG	HG	ME	HG	0.8
9	LO	FL	HG	ME	LO	ME	ME	0.8
10	LO	FL	HG	LO	ME	ME	LO	0.8
.								
.								
.								
71								

Table 8. Variables of the examples given as presented in Fig.1.

variable	question 1	2	3	4	5
PO	50000	250000	10000	5000	200000
TT	46	72	90	55	88
SW	4.8	3.0	1.5	5.0	3.8
TE	260	340	140	400	300
PR	20000	101	500	3000	1700
MA	2300	8000	10000	5000	6000

Table 9. Results of analysis.

question	property loss/ million dollars	fatalities / persons
1	45.80	6
2	-	-
3	10.00	1
4	17.28	2
5	-	-

Table 10. Fuzzified values of variables in questions 2 and 5 given as presented in Fig. 1.

variable	memb. function	2. question fuzzification degree	5. question fuzzification degree		
		1.	1.	2.	3.
PO	a	245000	196000	190000	180000
	b	250000	200000	200000	200000
	c	250000	200000	200000	200000
	d	255000	200000	210000	220000
TT	a	70.5	86.2	83.6	79.2
	b	72.0	88.0	88.0	88.0
	c	72.0	88.0	88.0	88.0
	d	73.5	89.8	92.4	96.8
SW	a	2.94	3.72	3.61	3.41
	b	3.00	3.80	3.80	3.80
	c	3.00	3.80	3.80	3.80
	d	3.06	3.90	3.99	4.18
TE	a	333.2	294.0	285.0	270.0
	b	340.0	300.0	300.0	300.0
	c	340.0	300.0	300.0	300.0
	d	346.8	306.0	315.0	330.0
PR	a	98.98	1666	1615	1530
	b	101.00	1700	1700	1700
	c	101.00	1700	1700	1700
	d	121.20	1734	1785	1870
MA	a	7600	5700	5400	4800
	b	8000	6000	6000	6000
	c	8000	6000	6000	6000
	d	8400	6300	6600	7200

Table 11. Results from fuzzification of questions 2 and 5.

question	fuzzification degree	property loss / million dollars	fatalities / persons
2	1.	5.55	6
5	1.	-	-
5	2.	-	-
5	3.	-	-