Forest Fire Prediction using Fuzzy Prototypical Knowledge Discovery.

José Angel Olivas, Francisco Pascual Romero

Escuela Superior de Informática.
Universidad de Castilla-La Mancha
Ronda de Calatrava 5, 13071 – Ciudad Real. Spain
e-mail: {jaolivas,fpromero}@inf-cr.uclm.es

ABSTRACT

An application of Zadeh's prototype theory in the Knowledge Acquisition process, is presented here, and as a practical example, to define a method for predicting the evolution of the forest fire occurrence-danger rate in INCEND-IA: A KBS for prediction and decision support in fighting against forest fires. This method then allows us to interpret any real cyclical situation using a previously discovered paradigm and define the current period. The FPKD (Fuzzy Prototypical Knowledge Discovery) is presented as a mechanism with the aim of generating Prototypes of Data (A new set of data sufficiently representative to be able to summarize or assimilate the behavior of any of the remaining data); but the concept of prototype is a fuzzy concept and Zadeh's Theory provides an appropriate framework for its application. Data Mining techniques have been used (decision trees, time series, clustering...). Thus, it is possible to calculate the grade of compatibility of a real situation with the prototypes and define the current period using these affinity values, with the objective of predicting the evolution of the following days.

Keywords: Artificial Intelligence, Data Mining, Prediction, Fuzzy Logic.

1 Introduction: Knowledge Acquisition, Data Mining and Fuzzy Prototypes.

The Knowledge Acquisition consists of the harvesting of the information necessary to construct a Knowledge Based System from any possible source. This information can be constituted by data, news, human knowledge, etc. The necessary information may appear in multiple forms: *Manuals and books*, *Formal Documentation*, *Informal Documentation*, *Internal registries, Presentations*, *Publications, Research*, *Visits, Human Knowledge.* The Knowledge Engineer must constantly control what information he needs, to what extent, in what subjects and what techniques he must use to acquire that specific knowledge and other factors. He is often tempted to improvise, which frequently causes negative results and lack of rigor.

For these reasons, three major blocks are basically contemplated: 1.- Evaluation of viability, definition of the problem, first meetings. 2.- Extraction of knowledge of the documentation (including data bases and other documentary sources previously mentioned). 3.- Educing of knowledge of the expert, managers and users.

In a more precise way, the objectives of Knowledge Acquisition are the following:

1.- General understanding of the task and functional structure of the KBS. 2.- Characterization of the reasoning process of the experts and the steps they normally use in the resolution of problems. 3.- Obtaining the necessary data to solve a particular problem, with the values that can take. 4.- Development of a conceptual model.

In addition to the usual methods of Knowledge Acquisition: Automatic methods, from examples, induction; Structural text analysis; Interviews (Open or structured); Observation of habitual tasks;

Classification of concepts; Questionnaires; Analysis of protocols; Repertory Grids, Techniques for educing in group; Delphi Method... The use of Data Mining helps us to handle great volumes of data, for the purpose of providing useful information and knowledge from them.

1.1 The KDD Process to extract useful knowledge of great volumes of data.

At the moment, our ability to analyze and to understand great sets of data is far below the capacity to store it. A new generation of techniques and computational tools becomes necessary for the extraction of useful knowledge, because a fast growth of the volume of data occurs generally. These techniques and tools are the subject of a new field of investigation called KDD: *Knowledge Discovery in Databases and Data Mining* [Fayyad, 96a].

In this research, the term KDD represents the entire discovery process of useful knowledge from data and Data Mining represents one more step in this process (application of specific algorithms to extract models of the data). Other steps such as preparation, selection and data-cleaning, incorporation from appropriate previous expert knowledge and interpretation of the results are contemplated. Therefore, the KDD takes and contributes theories, algorithms and methods of fields such as data bases, machine learning, pattern recognition, statistics, artificial intelligence and approximate reasoning and Knowledge Acquisition in Expert Systems [Fayyad, 96b].

1.2 Fuzzy Prototypes and FPKD process

1.2.1 Concepts and Prototypes

Fayyad and his collaborators define the KDD process as “the non-trivial process to identify valid, new, potentially useful and comprehensible patterns in data”, using multidisciplinary techniques and observing the manner in which they act together. The term *process* implies that there are several steps, such as the preparation of data or the search for patterns. The term *pattern* (in this study it will be called *prototype of data*) refers to a new set of data, along with a description and a model applicable to this set of data. The prototypes of data discovered must be valid for new data with some degree of certainty. It is necessary for these patterns to be new, and potentially useful. Finally, these patterns must be comprehensible, if not immediately, at least after a postprocessing. This definition implies that we have to consider measures of how good the prototypes of data are. In many cases it is possible to define measures of certainty (capacity of classification of new data) or utility (quality of the predictions on the basis of theses prototypes of data).

Taking the prototype theory from the cognitive psychology as a reference, a single representation of growth of the occurrence-danger rate of forest fires could be seen as prototypical. However, in a previous approximation at the Knowledge Acquisition process we were able to observe that this representation excessively simplifies the behavioral guidelines of the experts. When a technician is confronted with a real situation he handles a range of prototypes determined by a series of factors and must decide which type of evolution of the occurrence-danger rate is to be expected. Therefore, the prototype “Evolution of the occurrence-danger rate” is not unique[Zadeh, 82].

1.2.2 The Process

The stages of the modified KDD (from now on *Fuzzy Prototypical Knowledge Discovery*: FPKD) are the following ones: ***Selection****:* Applying the knowledge of the dominion and the excellent knowledge a priori, considering the objectives of the global process of FPKD a Target Data is created that will include selected sets of data or subgroups of excellent variables or examples. ***Pre-***

processing: Data Cleaning, elimination of noise, handling of empty fields, lost data, unknown or by defect values, evolution of data. Standard techniques of databases are applied. ***Transformation***: Reduction of the number of variables. Location of useful forms to express the data depending on the later use that are going away to give to them and on the objectives of the system. The expert knowledge and techniques of transformation and information in data bases are used. ***Data Mining***: Selection of the algorithms of Data Mining. Decisions about the model that is derived from the algorithm of chosen Data Mining (classification, summary of data, prediction). Search of interest patterns, as far as classification, rules of trees, regression, classification, dependency, heuristics, uncertainty.

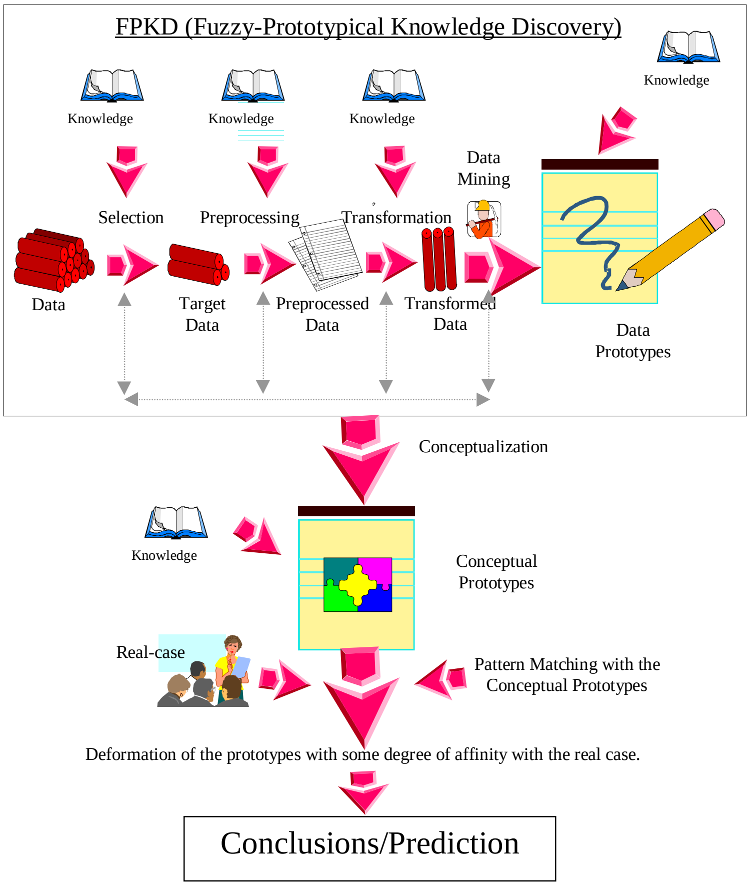

Figure 1. FPKD process.

1.3 Data Mining Algorithm and Techniques.

Within described process FPKD, the step of Data Mining acquires special relevance to determine the patterns of the observed data. The selection of the models to use has a fundamental component of experts knowledge, supervised by the Knowledge Engineer. In literature, for example [Berry, 96], it is described a great number of algorithms and techniques of pattern recognition, fuzzy logic and machine learning (for finding “regularities” in large data sets).

The algorithms of Data Mining consist of a specific mixture of three components [Fayyad, 96b]: The ***model***: Contains parameters that are determined from the data. Two relevant factors:

1. **The function**: *Regression, Dependency, Relation Analysis* between fields of the databases. The more excellent for our model, whose objective is the prediction or evaluation of a real case by

means of the comparison with prototypical situations, are the following: *Classification* Functions: Include a registry within one of several predefined classes. *Clustering* Functions: Include a registry within one of several classes (clusters), but unlike the classification, the classes are determined by the own data, by means of natural groups based on measures of affinity, similarity or probability. *Summary* Functions: They generate a compact description of a subgroup of data. A simple example like average can be used and the standard deviations for all the fields. Functions of *Analysis of sequences*: represent sequential patterns, temporary series. The objective is to generate the sequence of states of the process that tries to represent.

2. **The representation**: Classic models of representation are used, like rules, decision trees, linear and non-linear models (neural networks, genetic algorithms), methods based on examples (cases-based reasoning), bayesian networks, fuzzy models, etc. The representation model will determine the flexibility of representation of the data and interpretation of the model from the human point of view.

The preference criterion. It is that allows to select a model or set of parameters based on determined data. It could be a measure of goodness of the relation between the model and the data. Normally this criterion is explicit and quantitative in the search algorithm (for example, the criterion to find the parameters that maximize the probability of some observed data). Also usually there is a subjective implicit criterion (Knowledge Engineer), on what models consider initially.

The search algorithm. In Data Mining usually are used Artificial Intelligence techniques of optimization (for example, the *descent of the gradient, best-first search ...)*..

1.4 Algorithms for FPKD process.

An algorithm of Data Mining in FPKD process is an instantiation of the components model/preference/search, for example a model of classification based on a representation with a decision tree, a model of taxonomies based on the probability of the data using heuristics, etc. Each real problem requires a specific algorithm. The Knowledge Engineer must select the best *a priori* options, and if it is necessary, review and prove alternative models.

2 Forest Fires Prediction: Introduction to the problem

The problem of assigning and optimizing resources is a constant in the daily fight against forest fires in the Galician Area (Spain) and is due to the frequency and simultaneity of the fires together with the limited resources available. Thus, it seems necessary to predict the evolution of the forest fire occurrence-danger rate for a given area in the short and medium term. In order to satisfy this real prediction need, it is presented the following method as a sample of the research carried out for the production of a KBS prototype.

After an analysis of forest fires occurrence in the years 1991 and 1992, the experts have observed that the evolution of the occurrence-danger rate can be represented as a growing function of the sigmoidal type. This function is divided intro three sector starting the day after a rainy period, in which the number of forest fire occurrences has become zero. The growth pattern is repeated in a cyclic fashion after each rainy period but may suffer some modification due to specific factors which we will later on describe. It is understood as forest fire occurrence-danger rate the combination of the number and danger of the fires that took place. Figure 2 displays the representation of the growth pattern:

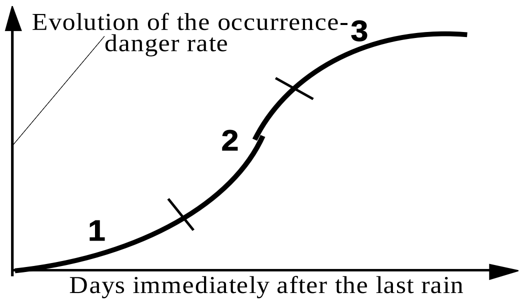

Figure 2. Representation of the growth pattern.

The first sector represents a slow growth of the occurrence-danger rate in the days immediately after a rainy period. The second sector depicts a high level of growth of the occurrence-danger rate, especially in terms of the number of fires. Finally, in the third sector the number of fires levels off but their danger level progressively increases.

3 FPKD: The search for fuzzy Prototypes.

3.1 Selection: The Target Data

It is taken as start set a relational database that contains about 12.000 forest fires registers, that toke place in Galicia (North-West of Spain) during the years 1991 and 1992. It is selected a subset of 3.204 forest fires corresponding to a smaller zone (Lugo, one of the four provinces of Galicia). Not relevant fields are also deleted.

3.2 Preprocessing: Data-Cleaning.

Once the 3.204 fires were extracted from the database, it should be proceed to their complete study. Some modifications in the design as well as in the content of table are done, because it presents irregularities that make impossible or very difficult the utilization of these data.

1.2.1 Missing Data and unknown values.

All the fields that are handled have a set of missing values, that is, the value or the values that are not considered as valid and there’s no default values. Furthermore, all the existing values are indispensable for the subsequent treatment.

3.2.2 Bad design of the databases.

The work database, not take into account the concepts of normalization as well as the essential characteristics of a good design, data were stored not thinking in a subsequent utilization. For example, to reflect the personal and mechanical resources that they were involved in the fire extinction. 20 fields of text type for human resources and other 20 for mechanical resources are used. These fields contains the names of the resources, but only are used a small part (not exceed 14 fields), that generate much lost space. The solution to this problem consists of eliminating these fields and introducing two new fields that summarize on the quantity of mechanical and human resources that have participated in the fight against each fire

3.2.3 Data Types.

The database contains the control of the dates and the hours in which the Fires have been begun, have been controlled and have been extinguished. These data are very important to control the temporary evolution of the fires and the statistic data obtainment on days, weeks... In the original database these fields were as simple chains of characters with a certain format, losing all its own semantics not been possible to calculate operations as comparisons, subtractions, distances. The solution is that the dates and hours may have been stored with some standard format.

3.3 Transformation.

3.3.1 Selection of relevant variables.

The relevant data that have been considered for the study of a cycle are the following: Date of beginning of the first fire of the cycle. Date of beginning of the last fire of the cycle. Total number of fires of the cycle. Number of fires, by type (Surface and materials that are burnt). Total wooded surface that it has been burnt during the cycle. Total level surface affected during the cycle. Quantity of human resources, mechanical and specialists that they have intervened in the fight against all the fires that have been started during the cycle.

3.3.2 Detection of cycles

The cycles are defined as sets of fires that have been produced in consecutive days (without mediating between them more than a day of difference). The detection process throws a result of 37 cycles in the 3.204 fires, including all the relevant information for their utilization in the following stages of the process, (see table 1).

A	B	C	D	E	F	G	H	I	J	K	L
0	2/01/91	2/01/91	1	0	0	1	0,01	0,5	3	1	1
1	15/01/91	15/01/91	1	1	0	0	0	0,001	1	1	0
2	24/01/91	30/01/91	9	2	4	3	0	0,272	10	10	2
3	23/02/91	25/02/91	12	1	8	3	0	0,395	14	13	2
4	4/03/91	4/03/91	1	0	0	1	0,008	0,03	3	1	1
5	11/03/91	11/03/91	1	0	1	0	0	0,03	1	1	0
6	20/03/91	20/03/91	3	1	0	2	0,032	0,13	2	3	1
7	29/03/91	2/04/91	45	17	19	9	0,213	1,209	31	60	45
8	9/04/91	11/04/91	20	6	10	4	0	0,406	20	23	14
9	14/04/91	24/04/91	48	24	14	10	0,392	0,824	64	71	40
10	27/04/91	27/04/91	1	1	0	0	0	0,01	0	1	1
11	8/05/91	29/05/91	113	59	18	36	0,99	3,641	162	185	126
12	1/06/91	8/06/91	7	4	2	1	0	0,076	9	7	5
13	11/06/91	19/06/91	19	15	0	4	0,169	0	24	30	30
14	22/06/91	6/07/91	57	39	10	8	0	1,841	72	102	87
15	10/07/91	10/09/91	554	375	94	85	0	14,922	744	1306	1196
16	14/09/91	25/09/91	87	58	24	5	0	1,335	94	196	198
17	2/10/91	2/10/91	1	1	0	0	0	0	0	1	0
18	6/10/91	7/10/91	8	6	2	0	0	0,109	10	20	6
19	1/11/91	1/11/91	1	0	1	0	0	0,03	1	2	0
20	24/11/91	24/11/91	1	0	1	0	0	0,02	4	2	1
21	29/11/91	30/11/91	2	1	1	0	0	0,03	2	2	0
22	5/12/91	7/12/91	16	10	5	1	0	0,396	9	25	7
23	14/12/91	16/12/91	6	1	4	1	0	0,245	7	14	3
24	21/12/91	8/01/92	213	98	78	37	63,876	434,756	147	383	136
25	16/01/92	20/01/92	9	8	1	0	1,01	2,71	7	12	4
26	23/01/92	11/02/92	424	233	112	79	206,3	1064,74	234	635	187
27	15/02/92	23/03/92	916	635	189	92	183,1	1074,52	366	1406	515
28	10/04/92	23/05/92	268	195	41	32	139,2	293,4	223	494	305
29	13/06/92	15/06/92	7	6	0	1	0,8	3,24	7	9	5
30	19/06/92	20/06/92	3	2	0	1	1,7	1,8	2	4	3
31	25/06/92	26/06/92	2	2	0	0	0	0,3	1	2	0
32	29/06/92	18/07/92	53	51	1	1	2,08	8,57	38	70	54
33	21/07/92	7/08/92	133	110	9	14	0	82,9	136	266	217
34	12/08/92	27/08/92	80	70	6	4	6,15	0	68	155	142
35	2/09/92	21/09/92	80	74	4	2	3,92	0	87	155	142
36	3/10/92	3/10/92	1	0	0	1	1	0,75	1	2	1

Table 1. The 37 cycles by date.

A: Identification Number of each cycle.
B: Start Date.
C: Finish Date.
D: Total number of fires.
E: Number of small fires (surface < 0.5 Ha).
F: Number of level surface fires.
G: Number of wooded surface fires.
H: Burnt wooded surface.
I: Burnt level surface.
J: Number of Specialists.
K: Number of Patrols.
L: Number of Mechanical resources (Planes, Helicopters, trucks...).

3.4. Clustering by Repertory Grids

In order to detect the relationships between the cycles, for obtaining those scarce medium or high progression of the occurrence-danger rate, is accomplished a hierarchical *clustering* process by *Repertory Grids*. The set of elements is constituted by the 37 cycles, and the constructions are the following seven:

Construction	*Values*
C1 Total number of fires.	[0 - 20] 1, [21- 30] 2, [31- 50] 3, [51 - 100] 4, [101 +] 5
C2 Burnt wooded surface.	[0 - 0,2] 1, [0,21 - 0,4] 2, [0,41 - 1] 3, [1,1 - 20] 4, [21 +] 5
C3 Burnt level surface.	[0- 1] 1, [1,1 - 5] 2, [5,1 - 100] 3, [101 - 1000] 4, [1001 +] 5
C4 Number of specialists.	[0- 10] 1, [11 - 20] 2, [21 - 50] 3, [51 - 100] 4, [101 +]
C5 Number of personnel patrols.	[0 - 10] 1, [11 - 50] 2, [51 - 100] 3, [101 - 300] 4, [301 +] 5
C6 Number of mechanical resources.	[0 - 10] 1, [11 - 30] 2, [31 - 50] 3, [51 - 100] 4, [101 +] 5
C7 Number of days of the cycle.	[0- 3] 1, [4 - 10] 2, [11 - 20] 3, [21 - 30] 4, [31 +] 5

Table 2. Constructions and Values.

Cyc.	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
C1	1	1	1	1	1	1	1	3	1	3	1	5	1	1	4	5	4	1	1
C2	1	1	1	1	1	1	1	2	1	2	1	3	1	1	1	1	1	1	1
C3	1	1	1	1	1	1	1	2	1	1	1	2	1	1	2	3	2	1	1
C4	1	1	1	2	1	1	1	3	2	4	1	5	1	3	4	5	4	1	1
C5	1	1	1	2	1	1	1	3	2	3	1	4	1	2	4	5	4	1	2
C6	1	1	1	1	1	1	1	3	2	3	1	5	1	2	4	5	5	1	1
C7	1	1	2	1	1	1	1	2	1	3	1	4	2	2	3	5	3	1	1

Cyc.	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
C1	1	1	1	1	1	5	1	5	5	5	1	1	1	4	5	4	4	1
C2	1	1	1	1	1	5	4	5	5	5	3	4	1	4	1	4	4	3
C3	1	1	1	1	1	4	2	5	5	4	2	2	1	3	3	1	1	1
C4	1	1	1	1	1	5	1	5	5	5	1	1	1	3	5	4	4	1
C5	1	1	1	2	2	5	2	5	5	5	1	1	1	3	4	4	4	1
C6	1	1	1	1	1	5	1	5	5	5	1	1	1	4	5	5	5	1
C7	1	1	1	1	1	3	2	3	5	5	1	1	1	3	3	3	3	1

Table 3. Repertory Grid.

To accomplish an analysis of *clusters* on elements, a proximity matrix is built that represents the different similarities of the elements, a matrix of 37 x 37 elements (the cycles) that above the diagonal represents the distances between the different cycles. Converting these values to percentages, a new table is created and the application of Repertory Grids Analysis Algorithm returns a graphic as a final result (fig. 4).

((((((((((0,1),4),5),6),10),17),19),20),21),31) -> 100% (**A**)

(2,12) -> 100% (**B**)

((18,22),23) -> 100% (**C**)

(34,35) -> 100% (**D**)

((((((((((0,1),4),5),6),10),17),19),20),21),31),(2,12)) -> 97%
(3,8) -> 97%
(14,16) -> 97%
(24,26) -> 97%
(27,28) -> 97%
(29,30) -> 97%

(((((((((((0,1),4),5),6),10),17),19),20),21),31),(2,12)),((18,22),23))-> 93%
((29,30),36) -> 93%

((((((((((((0,1),4),5),6),10),17),19),20),21),31),(2,12)),((18,22),23)),(3,8)) -> 90%
(7,9) -> 90%
(15,33) -> 90%
((24,26),(27,28)) -> 90%
(25,((29,30),36)) -> 90%

(((((((((((((0,1),4),5),6),10),17),19),20),21),31),(2,12)),((18,22),23)),(3,8)),13) -> 83%
(11,(15,33)) -> 83%
((14,16),(34,35)) -> 83%

((7,9),((14,16),(34,35))) -> 79%

((((((((((((((0,1),4),5),6),10),17),19),20),21),31),(2,12)),((18,22),23)),(3,8)),13),(25,((29,30),36))) -> 75%
(((7,9),((14,16),(34,35))),32) -> 75%
((11,(15,33)),((24,26),(27,28))) -> 75%

((((7,9),((14,16),(34,35))),32),((11,(15,33)),((24,26),(27,28)))) -> 61%

((((((((((((((0,1),4),5),6),10),17),19),20),21),31),(2,12)),((18,22),23)),(3,8)),13),(25,((29,30),36))),
((((7,9),((14,16),(34,35))),32),((11,(15,33)),((24,26),(27,28)))) -> 40%

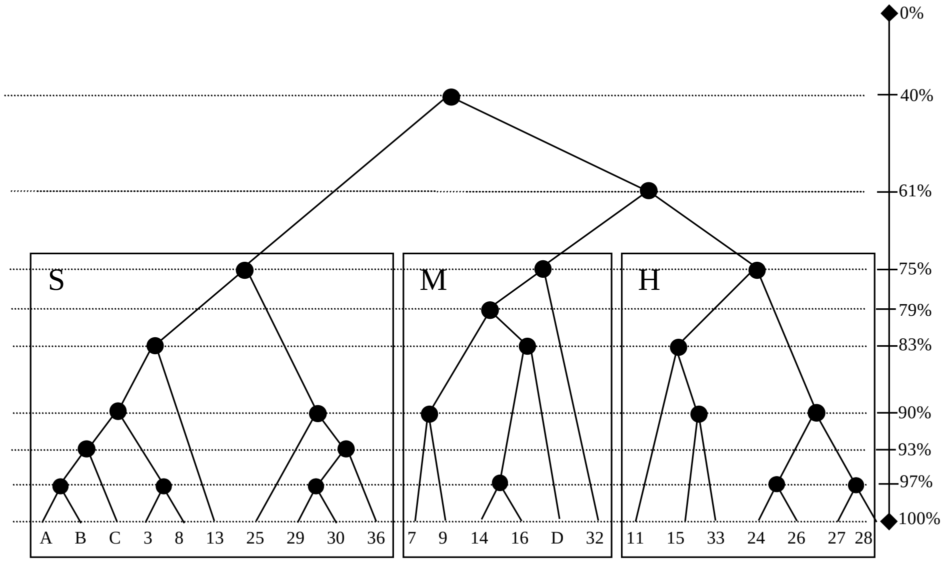

Figure 3. Clustering results (S: Scarce, M: Medium, H High progression cycles).

3.5 Classification of cycles

Taking into account this clustering process, the experts provide the necessary knowledge for defining a heuristic measure of danger-rate to evaluate each cycle:

Total number of fires / 100 + [(1* total wooded surface) + (0,5* total level Surface)] / 100

The experts agree: A cycle would be Scarcely progressive when this measure is below 0,4; between this value and 1 would be a Medium progressive cycle and from 1 onwards would be Highly progressive. Table 4 shows this classification:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	P
17	2/10/91	2/10/91	1	1	0	0	0	0	0	1	0	1	0,01	S
1	15/01/91	15/01/91	1	1	0	0	0	0,01	1	1	0	1	0,01	S
10	27/04/91	27/04/91	1	1	0	0	0	0,01	0	1	1	1	0,01	S
20	24/11/91	24/11/91	1	0	1	0	0	0,02	4	2	1	1	0,01	S
5	11/03/91	11/03/91	1	0	1	0	0	0,03	1	1	0	1	0,01	S
19	1/11/91	1/11/91	1	0	1	0	0	0,03	1	2	0	1	0,01	S
4	4/03/91	4/03/91	1	0	0	1	0,01	0,03	3	1	1	1	0,01	S
0	2/01/91	2/01/91	1	0	0	1	0,01	0,5	3	1	1	1	0,01	S
21	29/11/91	30/11/91	2	1	1	0	0	0,03	2	2	0	2	0,02	S
31	25/06/92	26/06/92	2	2	0	0	0	0,3	1	2	0	2	0,02	S
36	3/10/92	3/10/92	1	0	0	1	1	0,75	1	2	1	1	0,02	S
6	20/03/91	20/03/91	3	1	0	2	0,03	0,13	2	3	1	1	0,03	S
30	19/06/92	20/06/92	3	2	0	1	1,7	1,8	2	4	3	2	0,06	S
23	14/12/91	16/12/91	6	1	4	1	0	0,24	7	14	3	3	0,06	S
12	1/06/91	8/06/91	7	4	2	1	0	0,08	9	7	5	8	0,07	S
18	6/10/91	7/10/91	8	6	2	0	0	0,11	10	20	6	2	0,08	S
2	24/01/91	30/01/91	9	2	4	3	0	0,27	10	10	2	7	0,09	S
29	13/06/92	15/06/92	7	6	0	1	0,8	3,24	7	9	5	3	0,09	S
25	16/01/92	20/01/92	9	8	1	0	1,01	2,71	7	12	4	5	0,11	S
3	23/02/91	25/02/91	12	1	8	3	0	0,39	14	13	2	3	0,12	S
22	5/12/91	7/12/91	16	10	5	1	0	0,39	9	25	7	3	0,16	S
13	11/06/91	19/06/91	19	15	0	4	0,17	0	24	30	30	9	0,19	S
8	9/04/91	11/04/91	20	6	10	4	0	0,41	20	23	14	3	0,2	S
7	29/03/91	2/04/91	45	17	19	9	0,21	1,21	31	60	45	4	0,46	M
9	14/04/91	24/04/91	48	24	14	10	0,39	0,82	64	71	40	11	0,48	M
14	22/06/91	6/07/91	57	39	10	8	0	1,84	72	102	87	15	0,58	M
32	29/06/92	18/07/92	53	51	1	1	2,08	8,57	38	70	54	20	0,59	M
35	2/09/92	21/09/92	80	74	4	2	3,92	0	87	155	142	20	0,84	M
34	12/08/92	27/08/92	80	70	6	4	6,15	0	68	155	142	16	0,86	M
16	14/09/91	25/09/91	87	58	24	5	0	1,33	94	196	198	12	0,88	M
11	8/05/91	29/05/91	113	59	18	36	0,99	3,64	162	185	126	22	1,16	H
33	21/07/92	7/08/92	133	110	9	14	0	82,9	136	266	217	17	1,74	H
24	21/12/91	8/01/92	213	98	78	37	63,9	434	147	383	136	18	4,94	H
28	10/04/92	23/05/92	268	195	41	32	139	293	223	494	305	44	5,54	H
15	10/07/91	10/09/91	554	375	94	85	0	14,9	744	1306	1196	61	5,61	H
26	23/01/92	11/02/92	424	233	112	79	206	1064	234	635	187	19	11,6	H
27	15/02/92	23/03/92	916	635	189	92	183	1074	366	1406	515	39	16,4	H

Table 4. Cycles ordered by Prototypes.

The same that Table 1, plus:

M: Number of days of the cycle.
N: Heuristic Measure Value.
P: Prototype (S: Scarce, M: Medium, H High progression cycles).

3.6 Data Mining: Data Prototypes.

The selected algorithm for Data Mining process was *summarize functions*. The following table shows the parametric definition of the prototypes divided into sectors. This parameters will be modified taking into account the degree of affinity of a real situation with the prototypes. With the new modified prototype we will be able to predict the evolutions of the forest fires and optimize times, localization, etc. of the resources available for fighting against the fires.

	Scarce			Medium			High		
	S 1	**S 2**	**S 3**	**S 1**	**S 2**	**S 3**	**S 1**	**S 2**	**S 3**
Average of days	1	2	4	6	2	6	8	13	15
Average of fires/day	1	4	2	3	4	6	9	12	19
Minimun of fires/day	1	1	1	1	1	1	1	1	1
Maximun of fires/day	1	9	4	8	11	19	31	53	85
Number of Specialists/day	1	6	2	4	6	11	8	16	17
Number of Patrols/day	2	5	2	5	6	12	13	21	33
Number of Mech. Resources/day	1	3	1	4	5	11	5	13	18

Table 5. Prototypes “Scarce progression occurrence-danger rate“, “Medium progression“ and “High progression“.

4 FPKD: Conceptual Prototypes. Formal Representation.

As we have previously seen, the evolution of the occurrence danger rate is characterized by cycles which are initiated after each rainy period. The factors that *a priori* determine the progression of an occurrence danger rate from the rainy period to the next period can be reduced to three:

A. *Seasonal dryness*. Corresponds to a period of the year, keeping in mind that there can be a dry winter or a rainy summer. The user introduces the value at the beginning of each cycle: {High/0.95, Medium/0.45, Low/0.05}.

B. *Intensity of the last rainfall:* It is considered that influence of the intensity types is given by {Drizzle/0.9, Storm/0.5, Tempest/0.1}.

C. *Social Influence*: Represents the influence of the human factor on the evolution of the occurrence-danger rate (conflicts in the area, confrontations among neighbours, etc.) The experts determines this type of influence as {Much/0.85, Medium/0.45, Scarce/0.05}.

4.1 The Conceptual prototypes as fuzzy numbers

The prototypes have been represented as fuzzy numbers, which are going to allow us to obtain a grade o membership in the concept. For the sake of simplicity in the model, they have been represented as equal isosceles triangles. Therefore, in order to construct the prototypes (triangular fuzzy numbers) we only need to know their centerpoints (“center of the prototype”).

The process for determining the centers of the prototypes is the following:

1. The expert is presented with a battery of randomly chosen cases and is asked to point out the prototype under which each case falls. See example in table 4. the expert is provided with the information in the first three columns (the cases) and then must fill in the fourth column (the prototypes) using that information. In order to respect the experts language in the interview we only present him linguistic labels and not the averages (fifth column):

Seasonal dryness	Rainfall intensity	Social influence	Prog.-ocur. danger rate	Average
High	Drizzle	Much	High	0.900
High	Tempest	Much	Medium	0.633
Medium	Storm	Medium	Medium	0.467
Medium	Drizzle	Much	High	0.733
Low	Tempest	Medium	Scarce	0.200
Low	Drizzle	Much	Medium	0.600
High	Storm	Scarce	Medium	0.500
Medium	Tempest	Scarce	Scarce	0.200
Low	Storm	Low	Scarce	0.200

Table 6.

(ii) If some prototype ended up without any or with not enough cases, we performed the inverse process. That is, we asked the expert for exemplars of these prototypes. For example, since we have seen that there were few occurrences of the prototypes "high progression" and "scarce progression", we asked the expert to present cases of these prototypes (table 7).

Seasonal dryness	Rainfall intensity	Social influence	Prog.-ocur. danger rate	Average
High	Storm	Much	High	0.767

High	Drizzle	Medium	High	0.767
Medium	Tempest	Medium	Scarce	0.333

Table 7.

(iii) Consequently, the centers of the prototypes are the averages of the cases corresponding to each prototype. In our example:

"High progression" center: 0.791
"Medium progression" center: 0.550
"Scarce progression" center: 0.233

(iv) The last step is to represent the triangular fuzzy numbers using two factors:
- The center of each prototype.
- The distance from zero to the center of the first prototype which is used as the standard distance for constructing the base of the three triangles (figure 4).

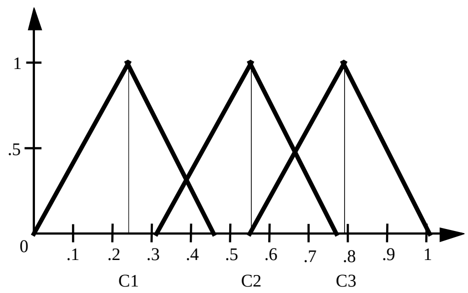

Figure 4.

Using another point of view, the second approach to solve this problem is the use of ID3 and CART algorithms, that in our practical cases get the same results.

5 Conclusions and future work.

Our objective has been to attempt the automation of an expert reasoning process, notwithstanding the fact that this reasoning cannot be formalized as a complete set of rules. The objective of this reasoning is the interpretation of a real situation. It seems correct to represent this interpretation process as a modification of prototypes to the real situation, just as we observed the expert do. This process was carried out in the following steps:

I. Conceptual representation:

(i) Determination of the concept we are going to study: "Progressive occurrence-danger rate" cycle or period that can be represented in a graph as a sigmoid divided into three sectors. The cycle starts and ends when there is a rainy period (null occurrence-danger rate). (ii) Determination of the factors that influence the evolution of the occurrence-danger rate (cycle factors). (iii) Determination of the three prototypes associated with the concept: "High", "Medium" and "Scarcely"; using FPKD.

II. Formal representation

(i) Calculate the center of each prototype asking the expert for examples and usin ID3 and CART algorithms. Represent the prototypes using triangular fuzzy numbers. (ii) Combine the factors of the real or current cycle in order to obtain their affinity with each of the prototypes. (iv) Define the current cycle beginning with the modification of the most similar prototypes, with a linear combination using de degree of affinity with the prototypes as weight values (Deformable Prototypes [Bremmerman, 76]. (v) Approximate the current cycle to the daily situation. This modification is achieved using rules that take into account the daily factors.

The method was validated using a database that contained two years of forest fire occurrences and asking the experts for the unknown information. From our point of view as well as that of the experts, we obtained positive results.

REFERENCES

[Adriaans, 96] Adriaans, P. W.; Zantinge, D.: *Data Mining*. Addison-Wesley, 1996.

[Agrawal, 96] Agrawal, R.; Mannila, H.; Srikant, R.; Toivonen, H.; Verkamo, A.: Fast Discovery of Association Rules. In *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, Cambridge MA, 1996.

[Berry, 96] Berry, M. J. A.; Linoff, G.: *Data Mining Techniques*. Wiley Computer Publishing. New York, 1996.

[Bremermann, 76] Bremermann, H.: *Pattern Recognition*. H. Bossel: Systems Theory in the Social Sciences. Birkhäuser Verlag, 1976, pp. 116 - 159.

[Fayyad, 96a] Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P.: The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*, November 1996/ Vol 39, Nº 11, pp. 27 – 34.

[Fayyad, 96b] Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P.; Uthurusamy, R. (Eds): *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, Cambridge MA, 1996.

[Olivas, 94] Olivas, J. A.; Sobrino, A.: An Application of Zadeh's Prototype Theory to the Prediction of Forest Fire in a Knowledge-based System. *Proceedings of the 5th. International IPMU* (Information Processing and Management of Uncertainty in Knowledge-based Systems), VOL. II, Paris, 1994, pp. 747 - 752.

[Zadeh, 82] Zadeh, L. A.: A note on prototype set theory and fuzzy sets. *Cognition* 12, 1982, pp. 291 - 297.

[Zadeh, 87] Zadeh, L. A.: *Fuzzy Sets and Applications* (Selected Papers, edited by R. R. Yager, S. Ovchinnikov, R. M. Tong, H. T. Nguyen), John Wiley, New York, 1987.